

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :  C07K 3/00, C12N 15/00	A1	(11) International Publication Number: WO 92/02536  (43) International Publication Date: 20 February 1992 (20.02.92)
---	----	--

(21) International Application Number: PCT/US91/05463

(22) International Filing Date: 1 August 1991 (01.08.91)

(30) Priority data:  
561,968 2 August 1990 (02.08.90) US

(71) Applicant: THE REGENTS OF THE UNIVERSITY OF COLORADO [US/US]; Dr. Chris Zafiratos, Vice President for Academic Affairs and Research, Campus Box 51, University of Colorado, Boulder, CO 80309 (US).

(72) Inventors: GOLD, Larry ; 955 - 8th Street, Boulder, CO 80302 (US). TUERK, Craig ; 3100 Dover Drive, Boulder, CO 80303 (US).

(74) Agents: SWANSON, Barry, J. et al.; 4582 S. Ulster St. Pkwy., Suite 403, Denver, CO 80237 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), NO, SE (European patent).

## Published

*With international search report.**Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

## (54) Title: SYSTEMATIC POLYPEPTIDE EVOLUTION BY REVERSE TRANSLATION

## (57) Abstract

The method of systematic polypeptide evolution reverse translation (SPERT) includes a candidate mixture of polypeptides having randomized amino acid sequence. Each member of the mixture is linked to an individualized mRNA which encodes the amino acid sequence of that polypeptide.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BC	Bulgaria	CN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU+	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TC	Togo
DE	Germany	MC	Monaco	US	United States of America

+ It is not yet known for which States of the former Soviet Union any designation of the Soviet Union has effect.

-1-

SYSTEMATIC POLYPEPTIDE EVOLUTION  
BY REVERSE TRANSLATION

5        This work was supported by grants from  
the United States Government funded through the  
National Institutes of Health. The U.S.  
Government has certain rights in this invention.

10      FIELD OF THE INVENTION

We describe herein novel high-affinity  
polypeptide ligands that specifically bind a  
desired target molecule. A method is presented  
for selecting a polypeptide ligand that  
15     specifically binds any desired target molecule.  
The method is termed SPERT, an acronym for  
Systematic Polypeptide Evolution by Reverse  
Translation. The method of the invention (SPERT)  
is useful to isolate a polypeptide ligand for a  
20     desired target molecule. The polypeptide products  
of the invention are useful for any purpose to  
which a binding reaction may be put, for example  
in assay methods, diagnostic procedures, cell  
sorting, as inhibitors of target molecule  
function, as probes, as sequestering agents and  
25     the like. In addition, polypeptide products of  
the invention can have catalytic activity. Target  
molecules include natural and synthetic polymers,  
including proteins, polysaccharides,  
30     glycoproteins, hormones, receptors and cell  
surfaces, nucleic acids, and small molecules such  
as drugs, metabolites, cofactors, transition state  
analogues and toxins.

-2-

BACKGROUND OF THE INVENTION

As translation of mRNA proceeds, stable complexes are formed. These complexes are made of ribosomes bound to mRNA with tRNA and nascent polypeptide encoded by the messenger RNA. Termed "ribosome complexes" herein, such complexes can be isolated by various known processes (Connolly and Gilmore (1986) J. Cell Biol. 103:2253; Perara et al. (1986) Science 232:348). Antigen-encoding mRNAs have been purified by taking advantage of the immunoreactivity of nascent polypeptides associated with ribosome complexes (Sambrook, J., Fritsch, E.F., Maniatis, T. Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, NY) (1989) ibid. sections 8.9-8.10). Such immunoreactive ribosome complexes can be immunoprecipitated from solution or separated by protein A column chromatography from non-reactive ribosome complexes (Schutz et al. (1977) Nuc. Acids Res. 4, 71; Shapiro and Young (1981) J. Biol. Chem. 256, 1495). Cyclical selection and amplification of RNAs with partitionable properties is now also possible. Historically, mRNA selection is closely tied to immunopurification of ribosome complexes, however, the partitioning of ribosome complexes according to the present invention is not restricted to immunoreactivity of the nascent polypeptides.

30 SUMMARY OF THE INVENTION

In its broadest aspect, the method of systematic polypeptide evolution by reverse translation (SPERT) includes a candidate mixture of polypeptides having a randomized amino acid

-3-

sequence. Each member of the mixture is linked to an individualized mRNA which encodes the amino acid sequence of that polypeptide. The candidate polypeptides are partitioned according to their 5 property of binding to a given desired target molecule. The partitioning is carried out in such a way, herein described, that each mRNA encoding a polypeptide is partitioned exactly together with that polypeptide. In this way each polypeptide is 10 partitioned together with the means for further amplifying it by an *in vitro* process. Ultimately, both the desired optimal polypeptide ligand of the desired target and the mRNA encoding the polypeptide are simultaneously selected, allowing 15 further synthesis of the selected polypeptide as desired, and further amplification of the coding sequence. It is therefore not necessary to analyze the amino acid sequence of the selected polypeptide (using protein chemistry) in order to 20 produce it in desired quantities.

Viewed another way, the invention is the selective evolution of a nucleic acid that encodes a polypeptide ligand of a desired target. The present method is therefore a selection based upon 25 coding properties available in a candidate nucleic acid mixture. In a previously filed application, U.S. Serial No. 07/536,428, filed June 11, 1990, incorporated herein by reference, the inventors herein have taught a method for selective 30 evolution of nucleic acids based upon binding properties of the nucleic acids themselves. The insight that cyclical selection and amplification can be a powerful tool for developing novel compounds when coupled with a partitioning system

-4-

is herein adapted to evolving specific coding nucleic acids, based on the partitioning properties of polypeptide ligands binding to target molecules.

5                  More specifically, the invention includes a method for making a polypeptide ligand of a desired target molecule which includes the following steps: First, synthesizing a mixture of translatable mRNA's, having certain sequence  
10                segments in common such as a ribosome binding site and a translation initiation codon and having a segment encoding a polypeptide at least part of which coding region is a randomized sequence.  
15                Second, employing the mRNA mixture in an in vitro translation system. Synthesis of nascent polypeptides ensues, each encoded by its own mRNA. At any time during translation, stable ribosome complexes can be isolated. It is preferred to isolate complexes in which translation has been  
20                stopped, or "stalled" by any of several known circumstances. Each isolated ribosome complex includes at least one ribosome, one nascent peptide and the coding mRNA which is now said to be translated mRNA. Although its chemical  
25                structure is unaltered, translated mRNA is bound to the ribosome complex in a different manner than it was bound prior to translation, as is known in the art. Third, the ribosome complexes are partitioned with respect to the binding of each  
30                nascent polypeptide to a desired target molecule. Some polypeptides bind weakly, some tightly, some not at all, with the target. The partitioning, however conducted, generally separates the mixture of ribosome complexes into ribosome complex-target

-5-

pairs and unbound complexes. The set of ribosome complex-target pairs is thereby enriched for those polypeptides (and, necessarily their coding mRNA's) that can bind to the target. Fourth, the 5 encoding mRNA's are separated from the complexes and amplified by conventional means for amplifying nucleic acids, such as reverse transcription and polymerase chain reaction (PCR). This amplification sets the stage for a subsequent 10 round of transcription, polypeptide synthesis and partitioning to further enrich for target-binding polypeptide ligands. These cycles can be reiterated as many times as desired, until a 15 desired binding affinity is achieved, or no further improvement in binding affinity is observed. The coding mRNA for any polypeptide selected in the foregoing manner can be cloned and sequenced, if desired. An individual polypeptide ligand can then be prepared in vivo from cloned 20 coding mRNA, or by chemical or enzymatic methods in vitro.

The present invention provides a class 25 of products which are polypeptides, each having a unique sequence, each of which has the property of binding specifically to a desired target compound or molecule. Each compound of the invention is a specific ligand of a given target molecule. The invention is based on the unique insight that cyclical selection and amplification of nucleic 30 acids can be applied to coding sequences by partitioning such coding sequences according to the binding affinities of the encoded polypeptides. In vitro evolutionary selection can therefore be applied for the first time to

-6-

polypeptides. Polypeptides have sufficient capacity for forming a variety of two- and three-dimensional structures and sufficient chemical versatility available within their monomers to act 5 as ligands (form specific binding pairs) with virtually any chemical compound, whether monomeric or polymeric. Molecules of any size can serve as targets. Most commonly, and preferably, for therapeutic applications, binding takes place in 10 aqueous solution at conditions of salt, temperature and pH near acceptable physiological limits. For other uses different binding conditions can be employed.

The invention also provides a method 15 which is generally applicable to make a polypeptide ligand for any desired target. The method involves selection from a mixture of candidates and step-wise iterations of structural improvement, using the same general selection theme, to achieve virtually any desired criterion 20 of binding affinity and selectivity.

While not bound by a theory of operation, SPERT is based on the inventors' insight that within a polypeptide mixture containing a large number of possible sequences 25 and structures there is a wide range of binding affinities for a given target. A polypeptide mixture comprising, for example a 10 amino acid randomized segment can have  $20^{10}$  candidate possibilities. Those which have the higher 30 affinity constants for the target are most likely to bind. After partitioning ribosome complexes, dissociation of mRNA and reverse transcription/amplification/ transcription, a

-7-

second polypeptide mixture is generated by translation, enriched for the higher binding affinity candidates. Additional rounds of SPERT progressively favor the best ligands until the 5 resulting polypeptide mixture is predominantly composed of only one or a few sequences. These can then be individually synthesized and tested for binding affinity as pure ligands. One cycle of SPERT effectively achieves reverse translation, 10 at least quantitatively.

Cycles of selection and amplification are repeated until a desired goal is achieved. In the most general case, selection/amplification is continued until no significant improvement in 15 binding strength is achieved on repetition of the cycle. The iterative selection/amplification method is sensitive enough to allow isolation of a single sequence variant in a mixture containing at least 65,000 sequence variants. The method could, 20 in practice, be used to sample about  $10^{18}$  different polypeptide species. There is no upper limit, in principle, to the number of different polypeptides which could be sampled, only a practical limit 25 dictated by the sizes of reaction vessels and other containers necessary to perform the method. The polypeptides of the test mixture include a randomized sequence portion as well as conserved sequences as desired for combining with other functional domains or to provide sufficient 30 polypeptide length to insure that the randomized sequence is accessible to the target in the ribosome complex. Amino acid sequence variants can be produced in a number of ways including chemical or enzymic synthesis of randomized

-8-

nucleic acid coding sequences. The variable sequence portion may contain fully or partially random sequence; it may also contain subportions of conserved sequence incorporated with randomized sequence. Sequence variation in coding nucleic acids can be introduced or increased by mutagenesis before or during the selection/amplification iterations.

In the case of a polymeric target, such as a protein, the ligand affinity can be increased by applying SPERT to a mixture of candidates comprising a first selected polypeptide sequence combined with a second randomized sequence. The sequence of the first selected ligand associated with binding or subportions thereof can be introduced into the randomized portion of the amino acid sequence of a second test mixture. The SPERT procedure is repeated with this second test mixture to isolate a second polypeptide ligand, having two sequences (one being the first polypeptide ligand) selected for binding to the target, which has increased binding strength or increased specificity of binding compared to the first polypeptide ligand isolated. The sequence of the second polypeptide ligand associated with binding to the target can then be introduced near the variable portion of the amino acid sequence after which cycles of SPERT results in a third polypeptide ligand. The third polypeptide ligand also contains the first and second ligand previously selected. These procedures can be repeated until a polypeptide ligand of a desired binding strength or a desired specificity of binding to the target molecule is achieved. The

-9-

process of iterative selection and combination of polypeptide sequence elements that bind to a selected target molecule is herein designated "walking," a term which implies the optimized binding to other accessible areas of a macromolecular target surface or cleft, starting from a first binding domain. Increasing the area of binding contact between ligand and target can increase the affinity constant of the binding reaction. These walking procedures are particularly useful for isolating novel polypeptides which are highly specific for binding to a particular target molecule.

A variant of the walking procedure employs a ligand termed "anchor" which is known to bind to the target molecule at a first binding domain (See Figure 8). This anchor molecule can in principle be any molecule that binds to the target molecule and which can be covalently linked directly or indirectly to a small bridge molecule for which a peptide binding sequence is known. When the target molecule is an enzyme, for example, the anchor molecule can be an inhibitor or substrate of that enzyme. The anchor can also be an antibody or antibody fragment specific for the target. The anchor molecule is covalently linked to the bridge molecule, chosen to bind an oligopeptide of known sequence. A test mixture of candidate polypeptides is then prepared which includes a randomized portion and includes also the known sequence that binds the bridging molecule. The bridging molecule binds the polypeptides to the target molecule in the vicinity of the anchor binding site. SPERT is

-10-

then applied to select polypeptides which bind a surface of the target molecule adjacent to the anchor binding site. Polypeptide ligands which bind to the target are isolated. Walking procedures as described above can then be applied to obtain polypeptide ligands with increased binding strength or increased specificity of binding to the target. Walking procedures could employ selections for binding to the anchor binding site itself or to another part of the target itself. This method is particularly useful to isolate polypeptide ligands which bind at a particular site within the target molecule. The anchor acts to ensure the isolation of polypeptide sequences which bind to the target molecule at or near the binding site of the anchor.

Screens, selections or assays to assess the effect of binding of a polypeptide ligand on the function of the target molecule can be readily combined with the SPERT methods. Specifically, screens for inhibition or activation of enzyme activity can be combined with the SPERT methods.

In more specific embodiments, the SPERT method provides a rapid means for isolating and identifying polypeptide ligands which bind to nucleic acids and proteins, including enzymes, receptors, antibodies, and glycoproteins.

In another aspect, the present invention provides a method for detecting the presence or absence of, and/or measuring the amount of a target molecule in a sample, which method employs a polypeptide ligand which can be isolated by the methods described herein. Detection of the target molecule is mediated by its binding to a

-11-

polypeptide ligand specific for that target molecule. The polypeptide ligand can be labeled, for example radiolabeled or enzyme linked, to allow qualitative or quantitative detection, 5 analogous to ELISA and RIA methods. The detection method is particularly useful for target molecules which are proteins. The method is more particularly useful for detecting proteins which are known to be only weakly antigenic, or for 10 which conventional monoclonal antibodies of a desired affinity are difficult to produce. Thus, polypeptide ligands of the present invention can be employed in diagnostics in a manner similar to conventional antibody-based diagnostics. One 15 advantage of polypeptide ligands over conventional antibodies in such detection methods and diagnostics is that polypeptides are capable of being readily synthesized in vitro or after cloning, since the method of the invention 20 concomitantly selects the means for amplification, e.g., coding nucleic acids, along with the ligand itself. Alternatively, the polypeptide can be chemically synthesized since its amino acid sequence can be ascertained readily from the 25 nucleotide sequence of its coding mRNA. A SPERT-generated polypeptide ligand need not be as large as an antibody molecule. Another advantage is that the entire SPERT process is carried out in vitro and does not require immunizing test 30 animals. Furthermore, the binding affinity of polypeptide ligands can be tailored to the user's needs. Compared to antibodies, SPERT-generated ligands have much greater versatility. Conventional antibodies are immunoglobulins,

-12-

which, although capable of a large repertoire of binding affinities, are nevertheless variations of a narrow amino acid sequence and structural theme. SPERT-generated polypeptide ligands, in contrast, 5 are unlimited as to structural type, and therefore have virtually unlimited potential for binding.

Polypeptide ligands of small molecule targets are useful as diagnostic assay reagents and have therapeutic uses as sequestering agents, 10 drug delivery vehicles and modifiers of hormone action. Catalytic polypeptides are selectable products of this invention. For example, by selecting for binding to transition state analogs of an enzyme catalyzed reaction, catalytic 15 polypeptides can be selected. Catalytic immunoglobulins have been developed by raising antibodies to transition state analogs (Schultz, P.C. (1989) Angew. Chem. Int. 2d Engl. 28:1283-1295; Schultz, P.G. (1989) Acc. Chem. Res. 22:287-294; Pollack, S.J. et al. (1989) Meth. Enzymol. 20 178:551-568).

In yet another aspect, the present invention provides a method for modifying the function of a target molecule using polypeptide 25 ligands which can be isolated by SPERT. Polypeptide ligands which bind to a target molecule are screened to select those which specifically modify function of the target molecule, for example to select inhibitors or 30 activators of the function of the target molecule. An amount of the selected polypeptide ligand which is effective for modifying the function of the target is combined with the target molecule to achieve the desired functional modification. This

-13-

method is particularly applicable to target molecules which are proteins. A particularly useful application of this method is to inhibit protein function, for example to inhibit receptor binding or to inhibit enzyme catalysis. In this case, an amount of the selected polypeptide molecule which is effective for target protein inhibition is combined with the target protein to achieve the desired inhibition.

The term "reverse translation" is used throughout as shorthand for the concept of information flow from polypeptide sequence to nucleic acid sequence. The phrase and shorthand make reference to the original and revised "central dogma" pronounced by Francis Crick many years ago. Crick understood and articulated the idea that either RNA or DNA could serve as a template for the synthesis of complementary nucleic acid sequences, and that chemically either RNA or DNA could serve as a template for the synthesis of both RNA and DNA. Crick noted that proteins, comprised of strings of amino acids, were templated by nucleic acid but could not serve themselves as a template for the synthesis of nucleic acids.

Most importantly, no simple chemistry is known that allows "reverse translation"; that was the basis nearly 25 years ago of Crick's adaptor hypothesis for using information in RNA to yield specified protein sequences during translation.

SPERT has at its center a form of reverse translation that does not conflict with Crick's postulates. While no process, no simple chemistry, is known that provides synthesis of a

-14-

nucleic acid containing a sequence specified by a polypeptide (whose sequence is unknown to the scientist at the time of reverse translation), SPERT provides a reliable mechanism for amplifying and using mRNAs that encode polypeptides of desired function but of unknown sequence.

Techniques for binding one or a few polypeptides to a selected target are known in the art, although binding of a small number of polypeptides from a randomized pool of polypeptides is of no value by itself. It is the concomitant selection in the ribosome complex of the mRNAs that encode those very polypeptides that provides a form of reverse translation because:

15                   1) the selected coding sequences can  
be amplified to yield large quantities of both DNA  
and RNA;

2) the newly made mRNA can be used for synthesizing polypeptides, now a smaller set than the original randomized mixture of polypeptides from which non-binding, or poorly-binding polypeptides have been removed, and;

3) the polypeptides held in ribosome complexes can be used for a subsequent round of SPERT.

Finally, "reverse translation" during SPERT does not yield a nucleic acid from only polypeptide sequence, but "reverse translation" does provide (through amplification techniques) net synthesis of the templates from which the desired polypeptide was synthesized. In principle a single molecule of polypeptide of the desired activity, along with a single template RNA in the translation complex, will lead to a nanomole or

-15-

even a micromole of nucleic acid corresponding to that polypeptide sequence. This net synthesis of nucleic acids based on the partitioning and activity of the desired polypeptide is an effective quantitative reverse translation that provides the materials for subsequent rounds of SPERT.

Also, the coding sequence can be used to deduce the amino acid sequence of a selected polypeptide. The polypeptide can then be synthesized by chemical methods, if desired.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a diagrammatic representation of steps in the process of the invention. The top panel depicts a double-stranded DNA template having a T7 promoter ("T7 PRO") and a segment of randomized sequence, represented as "nnn.....", preceded by a start codon, ATG. The initiation site of transcription and direction of transcription are shown as a vertical line labeled "+1" and an arrow, respectively. In vitro transcription creates mRNAs (2nd panel) which contain, from left to right, a ribosome binding site, a randomized sequence region, a 3' fixed sequence region, and a 3' primer annealing site. In vitro translation of this mixture gives rise to ribosome complexes with randomized nascent polypeptides (3rd panel). The ribosome complexes are subjected to selection for affinity of the nascent polypeptide and a desired target molecule (bottom panel). The encoding mRNAs of the partitioned complexes are purified and subjected to amplification, e.g., by reverse

-16-

transcription, PCR and transcription, to generate mRNAs for a second cycle of the process.

Figure 2 is a diagram showing expanded views of a ribosome complex. The top panel is a 5 ribosome complex as in the third panel of Figure 1. A cut-away view of the ribosome (2nd panel) shows 30-40 amino acids of the nascent polypeptide buried in the complex and unavailable for interaction with the solvent. The ribosome is depicted with two shades of gray to indicate inner 10 and outer regions. The nascent polypeptide is depicted as a thick white line extending vertically from a central tunnel (black) near the center of the ribosome. That portion inside the 15 ribosome is depicted as 30-40 amino acids in length. The carboxy-terminal end of the nascent polypeptide is shown connected to a peptidyl-tRNA (curly black line). The region bordered by a dotted line is expanded in the bottom panel 20 showing that the nascent polypeptide is covalently linked to a transfer RNA molecule which is hydrogen-bonded to the mRNA at a codon in the P-site.

Figure 3 is a diagram that represents 25 partitioning polypeptide ligands by direct immunoprecipitation. The top panel is a ribosome complex as in Figure 1. The center panel depicts several ribosome complexes where the nascent polypeptide is represented as a short, thick white line with hatching to indicate the segment of 30 randomized sequence. Molecules of a first antibody (immunoglobulin) are represented as inverted Y-shaped structures drawn with heavy, straight black lines. Interaction (binding) of a

-17-

nascent polypeptide with the epitope recognition site of an immunoglobulin is shown for two ribosome complexes. Nascent polypeptides are selected that have affinity for immunoglobulin molecules. The bottom panel shows addition of a second antibody (white inverted Y's) generally reactive to the first immunoglobulin resulting in an immunoprecipitate containing the selected ribosome complexes, shown as a cluster in the left half of the panel.

Figure 4 is a diagram showing partitioning of polypeptide ligands by indirect immunoprecipitation. The top panel shows a target protein which has an immunoreactive domain ("handle") and a target domain ("pan"). Three types of ribosome complexes are depicted in the second panel. Those with no affinity for the target protein are shown in white. Those with affinity for the "pan" are shown in light gray labeled with a "P" and shown with a bound target protein attached by the "pan" to the nascent peptide. Those with affinity for the "handle" are dark gray, labeled with an "H" and shown with a bound target protein attached by the "handle" to the nascent peptide. In the third panel, a first antibody (black lines) directed against the "handle" either displaces ligand associations of the "H" complexes or those complexes are unreactive. The first antisera form a sandwich with the "P" complexes made up of a ribosome complex associated with the target protein, through its "pan", and bound to the first immunoglobulin through the "handle". These "P" complexes are immunoprecipitated by second

-18-

antisera directed against the primary antisera, as shown in the bottom panel.

Figure 5 is a diagram showing selection of polypeptide ligands by membrane partitioning. The top panel shows a ribosome complex as in Figure 1. The middle panel shows ribosome complexes and membrane vesicles with membrane proteins. The membrane vesicles are depicted as a hatched band interrupted by hatched ovals that depict membrane proteins embedded in the membrane. In the middle panel, ribosome complexes are shown binding with membrane protein so that the nascent polypeptides having binding affinity for a membrane protein are partitioned. The bottom panel depicts three ribosome complexes bound to a membrane vesicle, forming a large complex which is separable from unbound ribosome complexes.

Figure 6 is a diagram showing partitioning of polypeptide ligands by affinity column chromatography. Ribosome complexes (top panel) are passed through a column containing insoluble support materials to which have been bonded target molecules. The middle panel is an expanded view of the column showing support materials (hatched circular segments) with attached target molecules (black bars) to which some ribosome complexes are bound. The bottom panel shows, enlarged, a single ribosome complex in which the nascent polypeptide (light shading) is bound to a target molecule which is attached to a column support bead (hatched). Ribosome complexes with high affinity to the target molecules are retained on the column and subsequently eluted to continue with SPERT.

-19-

Figure 7 is a diagram showing anchoring of a binding epitope and secondary ligand evolution. A molecule ("inhibitor") of known affinity for a target site on a protein is covalently linked to a "guide epitope". The guide epitope is any molecule for which there exists a peptide ligand, including a portion of a monoclonal antibody which contains an epitope recognition domain (Fab fragment). The mRNA encodes a reactive peptide sequence that binds the guide epitope, incorporated into the nascent polypeptide. The bottom panel depicts a ribosome complex having a nascent polypeptide that includes the reactive, guide binding, segment (shaded) and a randomized segment (unshaded). The ribosome complex is shown bound to the protein of interest by a binding interaction between the guide epitope and the reactive segment and by a secondary binding interaction between the randomized segment and a neighboring site on the target protein of interest. The randomized portion of the nascent polypeptide is free to evolve interactions with secondary sites on the target protein.

Figure 8 is a diagram which shows the DNA to be transcribed and the relationships of the oligonucleotides of Tables 1 and 2 in the DNA, prior to inserting the randomized sequence. The depicted structure constitutes a cassette for carrying out the transcription, translation, reverse transcription and PCR processes used in SPERT.

DETAILED DESCRIPTION OF THE INVENTION

The following terms are used herein

-20-

according to the definitions.

Polypeptide is used herein to denote any string of amino acid monomers capable of being synthesized by an in vitro translation system.

- 5       The term also embraces post-translational modifications introduced by chemical or enzyme-catalyzed reactions, as are known in the art. Such post-translational modifications can be introduced prior to partitioning, if desired.
- 10      Unless specified herein, all amino acids will be in the L-stereoisomeric form. Amino acid analogs can be employed instead of the 20 naturally-occurring amino acids. Any amino acid analog that is recognized by an aminoacyl-tRNA synthetase can
- 15      be employed. Several such analogs are known, including fluorophenylalanine, norleucine, azetidine-2-carboxylic acid, S-aminoethyl cysteine, 4-methyl tryptophan and the like.

- 20      Ligand means a polypeptide that binds another molecule (target). In a population of candidate polypeptides, a ligand is one which binds with greater affinity than that of the bulk population. In a candidate mixture there can exist more than one ligand for a given target.
- 25      The ligands can differ from one another in their binding affinities for the target molecule.

- 30      Candidate mixture is a mixture of nucleic acids and of polypeptides of differing sequence, from which to select a desired coding sequence or a desired ligand. The candidate mixture of nucleic acids serving as source of a candidate mixture of polypeptides can be in vitro transcription products of naturally-occurring nucleic acids or fragments thereof, chemically

-21-

synthesized nucleic acids, enzymatically synthesized nucleic acids or nucleic acids made by a combination of the foregoing techniques.

Target molecule means any compound of interest for which a ligand is desired. A target molecule can be a protein, fusion protein, peptide, enzyme, nucleic acid, nucleic acid binding protein, carbohydrate, polysaccharide, glycoprotein, hormone, receptor, receptor ligand, cell membrane component, antigen, antibody, virus, virus component, substrate, metabolite, transition state analog, cofactor, inhibitor, drug, controlled substance, dye, nutrient, growth factor, toxin, lipid, glycolipid, etc., without limitation.

Partitioning means any process whereby ribosome complexes bound to target molecules, termed complex-target pairs herein, can be separated from ribosome complexes not bound to target molecules. Partitioning can be accomplished by various methods known in the art. The only requirement is a means to separate complex-target pairs from unbound ribosome complexes. Columns which selectively bind complex-target pairs but not ribosome complexes, (or specifically retain ligand to an immobilized target) can be used for partitioning. A membrane or membrane fragment having the target on its surface can bind ligand-bearing ribosome complexes forming the basis of a partitioning based on particle size. The choice of partitioning method will depend on properties of the target and of the complex-target pairs and can be made according to principles and properties known to those of

-22-

ordinary skill in the art.

Amplifying means any process or combination of process steps that increases the amount or number of copies of a molecule or class of molecules. Amplifying coding mRNA molecules in the disclosed examples is carried out by a sequence of three reactions: making cDNA copies of selected mRNAs, using polymerase chain reaction to increase the copy number of each cDNA, and transcribing the cDNA copies to obtain an abundance of mRNA molecules having the same sequences as the selected mRNAs. Any reaction or combination of reactions known in the art can be used as appropriate, including direct DNA replication, direct mRNA amplification and the like, as will be recognized by those skilled in the art. The amplification method should result in the proportions of the amplified mixture being essentially representative of the proportions of different sequences in the mixture prior to amplification.

Specific binding is a term which is defined on a case-by-case basis. In the context of a given interaction between a given ligand and a given target, a binding interaction of ligand and target of higher affinity than that measured between the target and the candidate ligand mixture is observed. In order to compare binding affinities, the conditions of both binding reactions must be the same, and should be comparable to the conditions of the intended use. For the most accurate comparisons, measurements will be made that reflect the interaction between ligand as a whole and target as a whole. The

-23-

polypeptide ligands of the invention can be selected to be as specific as required, either by establishing selection conditions that demand the requisite specificity during SPERT, or by tailoring and modifying the ligands through "walking" and other modifications using iterations of SPERT.

Randomized is a term used to describe a segment of a nucleic acid or polypeptide having, in principle any possible sequence over a given length. Randomized nucleic acid sequences will be of various lengths, as desired, ranging from about twelve to more than 300 nucleotides. The chemical or enzymatic reactions by which random sequence segments are made may not yield mathematically random sequences due to unknown biases or nucleotide preferences that may exist. Redundancy of the genetic code, and biases in the tRNA content of an *in vitro* translation system can introduce additional bias in the translated amino acid sequences. Introducing a deliberate bias into a randomized coding region can reduce the bias of the resulting translated amino acid sequence. The term "randomized" is used instead of "random" to reflect the possibility of such deviations from non-ideality. In the techniques presently known, for example sequential chemical synthesis, large deviations are not known to occur.

A bias may be deliberately introduced into a randomized sequence, for example, by altering the molar ratios of precursor nucleoside (or deoxynucleoside) triphosphates of the synthesis reaction. A deliberate bias may be

-24-

desired, for example, to improve the randomness of amino acid sequence of translated polypeptides or to lower the frequency of appearance of certain amino acids.

- 5           For example, a randomized sequence biased for codons of the form ARN (where A is Adenine, R is Adenine or Guanine and N is any nucleotide) the most commonly encoded amino acids are basic (Arg, Asn, Lys) or polar (Ser).
- 10          Randomized sequences biased for codons of the form GRN are biased for acidic amino acids, Asp (GAU, GAC) and Glu (GAA, GAG), and Glycine (GGN). Randomized sequences in which U is never the 1st base in the triplet codon will lack termination signals and will not encode amino acids Phe, Tyr, Cys and Trp. By such strategies, randomized coding sequences can be biased for the type of structure likely to bind a given target. For example, polypeptide sequences biased for acidic amino acids can bind cationic target molecules more easily than completely random polypeptides.
- 20

- 25          Translatable mRNA is RNA which possesses all requisite sequences for translation in a conventional in vitro translation system. These include, improper orientation and sequence proximal to the 5' end of the RNA, a ribosome binding site and an initiation codon. In prokaryotes, as is known in the art, other codons, such as UUG and GUG can serve as initiation codons and encode methionine if properly spaced within a ribosome binding site.
- 30

Ribosome binding site means a nucleotide sequence in the mRNA which functions as a binding site for a ribosome in an in vitro translation

-25-

system.. The sequences which function as ribosome binding sites differ depending on whether the ribosomes are of procaryotic or eucaryotic origin, as is known in the art. In procaryotic systems, 5 the ribosome binding site is a short purine-rich region with a sequence such as GAGG or AGGA, usually located about 5 - 12 bases 5' to the initiation codon. The translation initiation codon is therefore usually located within 5 - 12 10 bases from the ribosome binding site in the 3' direction on the mRNA. These sequences are sometimes termed a Shine-Dalgarno sequence. The structures of ribosome binding sites and their proper placement to ensure correct initiation of 15 protein synthesis are well known in the art.

Initiation codon is a characteristic trinucleotide sequence AUG which encodes methionine and which encodes a first amino acid of an encoded polypeptide and also sets the codon 20 reading frame for the nucleotide sequence in the 3' direction from the initiation codon.

Ribosome complex is a macromolecular complex including at least one ribosome, attached mRNA molecule and, for each ribosome, a nascent 25 polypeptide attached via tRNA to the ribosome. The nascent polypeptide has an amino acid sequence encoded by the attached mRNA. Ribosome complexes are formed, as is known in the art, during protein synthesis. Ribosome complexes are stable if they 30 become stalled for any reason, for example, by depletion of release factor, lack of termination codon in the message, lack of a charged tRNA, etc., as known in the art. The mRNA together with attached ribosome(s) and nascent peptide(s) remain

-26-

stably bound and can be isolated together, using methods known in the art.

5           In vitro translation can be carried out using known systems. These well-known translation systems are the E. coli system, the wheat germ system, and the rabbit reticulocyte system. The latter is available commercially. The conditions for carrying out in vitro translations are well-known in the art, and various modifications, 10 adaptations and optimizations are available to those skilled in the art.

15           The combination of translatable mRNA encoding a polypeptide and in vitro translation system constitute amplifying means for amplifying the quantity of polypeptide encoded by the mRNA. The mRNA can itself be amplified using reverse transcription, PCR with appropriate primers and an RNA polymerase. The amplified mRNA can serve for 20 in vitro synthesis of desired quantities of the encoded polypeptide. As noted, supra, this process constitutes reverse translation.

25           The terms "ribosome" and "nascent peptide" have conventional meanings known in the art. The term "translated mRNA" simply refers to mRNA present in a ribosome complex, either wholly or partially translated.

30           Ribosome complex-target pairs are ribosome complexes of which the nascent polypeptide component is bound to a target molecule. The target molecule can be free in solution or bound to a solid support matrix.

Homology is used to compare the related uses of sequences. Percent amino acid sequence homology is measured by comparing sequences of

-27-

equal length position by position. The percent of those positions occupied by the same amino acid in two polypeptides is the percent sequence homology. Thus, given peptide ABCDE as a naturally-occurring  
5 comparison peptide, peptides ABCDX or ABXDE are 80% homologous but peptides ABXYZ, AXYZE and XYZDE are 40% homologous and peptides EDCBA, BDAEC, MNOPQ are non-homologous.

The SPERT method involves the  
10 combination of a selection of polypeptide ligands which bind to a target molecule, for example a protein, with amplification of those selected polypeptides via the attached mRNAs. Iterative cycling of the selection/ amplification steps  
15 allows selection of one or a small number of polypeptides which bind most strongly to the target from a pool which contains a very large number of nucleic acids and hence encoded polypeptides.

20 Cycling of the selection/amplification procedure is continued until a selected goal is achieved. For example, cycling can be continued until a desired level of binding of the polypeptides in the test mixture is achieved or  
25 until a minimum number of polypeptide components of the mixture is obtained (in the ultimate case until a single species remains in the test mixture). In many cases, it will be desired to continue cycling until no further improvement of binding is achieved. It may be the case that certain test mixtures of polypeptides show limited improvement in binding over background levels during cycling of the selection/ amplification.  
30 In such cases, the sequence and length variation

-28-

in the test mixture should be increased until improvements in binding are achieved. Anchoring protocols and/or walking techniques can be employed as well.

5            Specifically, the method requires the initial preparation of a test mixture of candidate polypeptides. A translatable mRNA mixture is prepared, each member of the mixture including in its nucleotide sequence a ribosome binding site, 10          an initiation codon and a randomized coding region. Preferably the individual mRNA's contain a randomized region flanked by sequences conserved in all nucleic acids in the mixture. The conserved regions are provided to facilitate 15          amplification of selected nucleic acids. Since there are many such sequences known in the art, the choice of sequence is one which those of ordinary skill in the art can make, having in mind the desired method of amplification. The 20          randomized coding region can have a fully or partially randomized sequence according to the desired translation product. Depending on the desired polypeptide structure, the coding portion of the nucleic acid can contain subportions that 25          are randomized, along with subportions which are held constant in all nucleic acid species in the mixture. For example, sequence regions known to code for amino acid sequences that bind, or have been selected for binding, to the target can be 30          integrated with randomized coding regions to achieve improved binding or improved specificity of binding. Sequence variability in the polypeptide test mixture can also be introduced or augmented by generating mutations in the coding

-29-

mRNA's during the selection/amplification process. In principle, the mRNA's employed in the test mixture can be any length as long as they can be amplified. The method of the present invention is most practically employed for selection from a large number of sequence variants. Thus, it is contemplated that the present method will preferably be employed to assess binding of polypeptide sequences ranging in length from about 5 four amino acids to any attainable size.

10

The randomized portion of the coding nucleic acids in the test mixture can be derived in a number of ways. For example, full or partial sequence randomization can be readily achieved by 15 direct chemical synthesis of the nucleic acid (or portions thereof) or by synthesis of a template from which the nucleic acid (or portions thereof) can be prepared by use of appropriate enzymes. Chemical synthesis provides the advantages of 20 being precisely controllable as to length and allowing individual randomization at each triplet position. A commercial DNA synthesizer can be used, either with an equivalent mixture of the four activated nucleotide substrates or with a 25 biased mixture. Alternatively, the synthesizer can be set up to provide a limited nucleotide selection at a given position, e.g., only A at the first triplet position. End addition, catalyzed by terminal transferase in the presence of 30 nonlimiting concentrations of all four nucleotide triphosphates can add a randomized sequence to a segment. Sequence variability in the coding nucleic acids can also be achieved by employing size-selected fragments of partially digested (or

-30-

otherwise cleaved) preparations of large, natural nucleic acids, such as genomic DNA preparations or cellular RNA preparations. In those cases in which randomized sequence is employed, it is not necessary (or possible from long randomized segments) that the test mixture contains all possible variant sequences. It will generally be preferred that the test mixture contain as large a number of possible sequence variants as is practical for selection, to insure that a maximum number of potential amino acid sequences of the translated polypeptide are identified. A randomized sequence of 60 nucleotides will contain a calculated  $10^{36}$  different candidate nucleic acid sequences which would encode  $10^{26}$  possible decapeptides. As a practical matter, it is possible to sample only about  $10^{18}$  polypeptide candidates in a single selection. Therefore, candidate mRNA mixtures that have randomized segments longer than 60 contain too many possible sequences for all to be sampled in one selection. Many epitopes recognized by antibodies are only 5-10 amino acids in length. It is not necessary to sample all possible sequences of a candidate mixture to select a polypeptide ligand of the invention. It is basic to the method that the coding nucleic acids of the test mixture are capable of being amplified. Thus, it is preferred that any conserved regions employed in the test nucleic acids do not contain sequences which interfere with amplification.

The complex of a ribosome, mRNA, and nascent polypeptide attached to a tRNA in the P-site of the ribosome is very stable. Release of

-31-

the nascent peptide from the complex and of the mRNA from the ribosome requires protein release factors. Release factor recognition requires the positioning of the stop codons of the mRNA in the A-site of the ribosome. In the absence of a stop codon or release factor the dissociation of the translation complex from mRNA is very slow. The addition of the antibiotics cycloheximide (eukaryotic systems) and chloramphenicol (prokaryotic system) further stabilizes the complexes so that extensive manipulations like column chromatography and gradient centrifugation can be performed.

For this invention, a ribosome is preferably paused at the end of a coding sequence on a mRNA with the encoded nascent polypeptide available for partitioning of the complex. There are a number of ways in which this can be accomplished. Because stop codons are essential for release factor action, a translating ribosome that does not encounter any stop codons will proceed to the end of a mRNA and stall at the 3' end (Connolly and Gilmore, supra). In vitro translation systems which have been depleted of release factor (by immunoinactivation or mutation) will result in the stalling of translation complexes at stop codons. Removal of GTP, the use of non-hydrolyzable analogues, and the use of certain antibiotics will also stall translational complexes. The timed addition of these exogenous factors to a synchronous in vitro translation reaction can produce predictable sizes of nascent polypeptide for the successful partitioning of the translational complex. In some organisms there

-32-

exist temperature-sensitive tRNA synthetase mutants. Another way of stalling translational complexes at defined sites is to include at the 3' end of the coding region a stretch of sense codons which are recognized by a single species of tRNA for which there exists a conditional tRNA synthetase mutant. In vitro translation reactions done from extracts of such mutants under the restrictive condition will result in stalled complexes at the stretch of sense codons for that particular tRNA.

It will be understood that it is not necessary to stall or pause the translation process to obtain partitionable ribosome complexes. Stable complexes can be isolated at any time during active translation. It is advantageous to isolate actively translating ribosome complexes when it is desired to vary the length of the randomized segment, e.g., to test the effects of polypeptide length on binding efficacy. Ribosome complexes isolated during active translation constitute a population of nascent peptides of varied length. By synchronously initiating translation and isolating ribosome complexes at various times thereafter, the effects of increasing polypeptide length can be compared.

Polymerase chain reaction (PCR) is an exemplary method for amplifying nucleic acids. Descriptions of PCR methods are found, for example in Saiki et al. (1985) Science 230:1350-1354; Saiki et al. (1986) Nature 324:163-166; Scharf et al. (1986) Science 233:1076-1078; Innis et al. (1988) Proc. Natl. Acad. Sci. 85:9436-9440; and in

-33-

U.S. Patent 4,683,195 (Mullis et al.) and U.S. Patent 4,683,202 (Mullis et al.). In its basic form, PCR amplification involves repeated cycles of replication of a desired single-stranded DNA (or cDNA copy of an RNA) employing specific oligonucleotide primers complementary to the 3' ends of both strands, primer extension with a DNA polymerase, and DNA denaturation. Products generated by extension from one primer serve as templates for extension from the other primer. A related amplification method described in PCT published application WO 89/01050 (Burg et al.) requires the presence or introduction of a promoter sequence upstream of the sequence to be amplified, to give a double-stranded intermediate. Multiple RNA copies of the double-stranded promoter-containing intermediate are then produced using RNA polymerase. The resultant RNA copies are treated with reverse transcriptase to produce additional double-stranded promoter containing intermediates which can then be subject to another round of amplification with RNA polymerase. Alternative methods of amplification include among others cloning of selected DNAs or cDNA copies of selected RNAs into an appropriate vector and introduction of that vector into a host organism where the vector and the cloned DNAs are replicated and thus amplified (Guatelli, J.C. et al. (1990) Proc. Natl. Acad. Sci. 87:1874). In general, any means that will allow faithful, efficient amplification of selected nucleic acid sequences can be employed in the method of the present invention. It is only necessary that the proportionate representations of sequences after

-34-

amplification reflect the relative proportions of sequences in the mixture before amplification.

Specific embodiments of the present invention for amplifying RNAs are based on Innis et al. (1988) supra. The RNA molecules in the test mixture are designed to contain a sequence transcribed from a T7 promoter in their 5' portions. Full-length cDNA copies of selected mRNA molecules are made using reverse transcriptase primed with an oligomer complementary to the 3' sequences of the selected RNAs. The resultant cDNAs are amplified by Tag DNA polymerase chain extension, employing a primer containing the T7 promoter sequence as well as a sequence complementary to the conserved 5' end of the selected RNAs. Double-stranded products of this amplification process are then transcribed in vitro. Transcripts are used in the next selection/amplification cycle. The method can optionally include appropriate nucleic acid purification steps.

In general, any protocol which will allow selection of polypeptides based on their ability to bind specifically to another molecule, i.e., a protein or any target molecule, can be employed in the method of the present invention. It is only necessary that the ribosome complexes be partitioned without disruption such that the selected coding mRNA's are capable of being amplified. For example, in a column binding selection in which a test mixture of ribosome complexes bearing nascent randomized polypeptide is passed over a column of immobilized target molecules, the complexes bearing polypeptide

-35-

ligands of the target are retained and the non-target binding complexes are eluted from the column with appropriate buffer. A wide variety of affinity chromatography techniques, including support matrices and coupling reactions is available for application of a column partitioning system. Target binding polypeptides together with mRNA's encoding each remain bound to the column. The relative concentrations of protein to test polypeptides in the incubated mixture influences the strength of binding that is selected for. When polypeptide is in excess, competition for available binding sites occurs and those polypeptides which bind most strongly are selected. Conversely, when an excess of target is employed, it is expected that any polypeptide that binds to the target will be selected. The relative concentrations of target to polypeptide employed to achieve the desired selection will depend on the type of target, the strength of the binding interaction and the level of any background binding that is present. The relative concentrations needed to achieve the desired partitioning result can be readily determined empirically without undue experimentation. Similarly, it may be necessary to optimize the column elution procedure to minimize background binding. Again such optimization of the elution procedures is within the skill of the ordinary artisan.

An unexpected feature of the invention is the fact that the polypeptide ligand need not be elutable from the target to be selectable. This is because it is the mRNA that is recovered

-36-

for further amplification or cloning, not the polypeptide itself. It is known that some affinity columns can bind the most avid ligands so tightly as to be very difficult to elute. However 5 the method of the invention can be successfully practiced to yield avid ligands, even covalent binding ligands. The ribosome complexes can be disrupted by denaturing agents such as urea or sodium dodecyl sulfate without affecting the 10 integrity of the mRNA. The mRNA's of selected ligands are amplified, as described elsewhere herein, to yield a mixture of coding sequences enriched for those that encode polypeptide ligands of the desired target, including ligands that bind 15 tightly, irreversibly or covalently.

Immunoreactivity of nascent polypeptides on ribosome complexes can be used to purify the encoding mRNAs. In the simplest application, 20 ribosome complexes are purified from cells in the presence of inhibitors such as chloramphenicol or cycloheximide which stall translational complexes on mRNA. Binding of antibodies which recognize the epitope of interest followed by binding antibodies which recognize those antibodies 25 results in immunoprecipitation of the ribosome complexes containing the mRNAs which encode the epitope. The background of mRNAs which do not encode the epitope of interest but are trapped by the immunoprecipitated complex can be lowered by 30 using purified IgGs against the epitope followed by purification of the immunoreactive ribosomes on a protein A column. (IgGs are one class of the soluble immunoglobulins which compose antisera. Protein A is derived from Staphylococcus aureus

-37-

and has a high affinity for IgGs. Protein A binding does not interfere with epitope recognition.)

These procedures for immunoprecipitation to partition ribosome complexes can be used in a variety of modifications to partition the translational complexes in SPERT. One such modification is termed "panhandling" (See Figure 4). A protein is composed of an immunoreactive domain for which known antibody exists, and a separate target domain for which one wishes to evolve protein ligands. Ribosome complexes which interact with the target domain (the "pan") via their nascent polypeptides will be immunoprecipitated upon binding antibodies which recognize the immunoreactive domain (the "handle"). This modification is especially useful for developing polypeptide ligands against a segment of a fusion protein in which the amino terminus is the fragment of a common protein ( $\beta$ -galactosidase, for example) and the carboxyl-terminal portion is the protein of interest. It will also be useful for the development of polypeptide ligands which recognize immunoresistant domains of a protein which has an immuno-dominant domain for which polyclonal sera is available. Where immunoprecipitation is employed, it will be advantageous to discard any ribosome complexes that react directly with the antibodies, prior to selection.

Alternative partitioning protocols for separating polypeptides bound to targets, particularly proteins, are available to the art. For example, binding and partitioning can be

-38-

achieved by immunoprecipitation of the test ribosome complex mixture and passing the immune complexes through a protein A affinity column which retains the immune reactive polypeptide-containing complexes as the column. Those mRNA's that encode a polypeptide that binds to the target antibody will be retained on the column as part of the ribosome complex and unbound coding mRNA's can be washed from the column.

Effective partitioning can be carried out with pure or impure target preparations. In cases where target preparations are impure, selectivity can be enhanced by strategies that enhance the binding of ligands to the desired target, or which specifically elute desired ligands or prevent their binding. The latter approach is subtractive. A known ligand can block binding of any polypeptide that can bind the target so that the desired polypeptide is partitioned by elution and unwanted polypeptides are retained on the column.

Optionally, chemical or enzymic modifications of the polypeptide can be introduced post-translationally. The process for making such modifications should not disrupt the ribosome complexes. An important type of post-translational modification is oxidation to form disulfides in sequences that contain two or more cysteines. Particularly for small polypeptides, disulfide bonds are especially advantageous to lock in a desired conformational state so that a rigid structure having high specificity and binding affinity for a target can be achieved. (See, e.g., Olivera, B. M., et al. (1990) Science

-39-

249:257-263.

Other forms of post-translational structure modifications include introducing factors that non-covalently influence tertiary structure of the nascent polypeptide. In particular, metal ions such as Ca<sup>++</sup>, Mg<sup>++</sup>, Mn<sup>++</sup>, Zn<sup>++</sup>, Fe<sup>++</sup>, Fe<sup>+++</sup>, and Mo<sup>6</sup> can affect polypeptide folding configuration by forming coordination complexes with amino acid side chains. Similarly organic compounds such as nicotinamide nucleotides, flavine nucleotides, porphyrins, thiamine phosphates, serotonin, and the like, including inhibitors, agonists and antagonists of known biological functions, can interact with the nascent polypeptide to modify its 3-dimensional folded configuration. As thus modified, the nascent polypeptide can exhibit different binding properties than an unmodified polypeptide. The use of such configurational modifiers enhances the range of potential binding activities of any candidate mixture of polypeptides. Also, it affords a means for selecting polypeptides having conditionally reversible functions, i.e., capable of being functionally "off" or "on", depending on the presence or absence of the modifier. Configurational modifiers need not be naturally-occurring compounds. The use of such modifiers during partitioning is only limited by the need to maintain stability of the ribosome complexes. Modifiers which disrupt ribosome complexes or which degrade the coding mRNA or nascent polypeptide should be avoided. A modifier can be included in the buffer or medium during partitioning. Alternatively, SPERT itself can be

-40-

used to pre-select polypeptides which bind the modifier as a target after which the candidate mixture of selected modifier-binding polypeptides can be further selected, via SPERT, for binding the ultimate target.

5 Sequence variation in the test coding mRNA mixture can be achieved or increased by mutation. For example, a procedure has been described for efficiently mutagenizing nucleic acid sequences during PCR amplification (Leung et al. 1989). This method or functionally equivalent methods can optionally be combined with amplification procedures in the present invention.

10 Alternatively, conventional methods of DNA mutagenesis can be incorporated into the nucleic acid amplification procedure. Applicable mutagenesis procedures include, among others, chemically induced mutagenesis and oligonucleotide site-directed mutagenesis.

15 The starting mRNA mixture is not limited to sequences synthesized de novo. In particular, SPERT can be used to modify the function of existing proteins. A segment of the natural sequence is replaced by a corresponding segment of randomized sequence in the mRNA that encodes the protein. Since many known proteins belong to families having some sequences conserved and others varied, the logical approach is to replace the variable (or hypervariable) regions with randomized sequence, to maximize the chance of altering function. The proper choice of partitioning conditions, as will be apparent to those skilled in the art, results in selection for the desired functional variant. In this way,

-41-

modifications, alterations and improvements on known proteins can be achieved.

In order to proceed to the amplification step, coding nucleic acids must be released from the target-bound ribosome complexes after partitioning. This process must be done without chemical degradation of the coding mRNA's and must result in amplifiable nucleic acids. In a specific embodiment, selected coding RNA molecules are eluted from a column using a high ionic strength buffer or other eluant capable of disrupting the ligand-target bond. Alternatively, the ribosome can be denatured such that the mRNA is eluted. The coding mRNA can be removed from ribosome complexes or from ribosome complex-target pairs by phenol extraction or by phenol combined with a protein denaturing agent such as 7M urea. Although ribosomal RNA is also extracted, subsequent amplification is selective for the mRNA's because the primers used for cDNA synthesis and PCR amplification are complementary only to a conserved sequence in the mRNA's and not to ribosomal RNA.

As the translation of randomized mRNAs proceeds during the SPERT protocol, the growing polypeptide makes its way from the peptidyl transferase site within the large ribosome subunit toward the cytoplasmic solvent. The peptidyl transferase site is an intrinsic activity of the large ribosome subunit from all organisms; that site has been defined functionally but its precise location within the ribosome is unknown. However, the distance between that site and the cytoplasmic solvent also is known to be about 30 to 40 amino

-42-

acids in length.

For optimal effectiveness in SPERT, the random portion of the nascent polypeptide (whose properties are selected during the procedure) should be "outside" the ribosome in order for partitioning of the ribosome complex to fully utilize the properties of the randomized polypeptide. A C-terminal trailer sequence is preferably incorporated into the translated polypeptide to insure that the randomized sequence is fully exposed after translation. From the work of Smith et al, (PNAS, 75:5922, 1978) and Malkin and Rich (J. Mol. Biol., 26:329, 1967) for both prokaryotes and eukaryotes: about 30 to 40 amino acid residues remain within the ribosome during translation. Furthermore, if the amino-terminus of a growing polypeptide contains a hydrophobic domain of about 20 amino acid residues, a nascent polypeptide of about 50 residues has been shown to be enough to allow the translation complex to interact with a membrane by hydrophobic interactions, see Kurzchalia et al, Nature 320:634, 1986). Thus, in preferred embodiments of SPERT, the randomized polypeptide will be encoded by randomized mRNA that is about 30-40 codons (that is, about 90-120 nucleotides) upstream from the codons at which the translation complex stalls. It will be understood that both longer and shorter C-terminal trailer sequences can be used effectively, and that SPERT, itself, can be used to determine optimum trailer length for a given partitioning system. The sequence of mRNA and encoded polypeptide in the C-terminal trailer can be designed to have any other desired

-43-

function, such as more stability in the translation complex, ease of in vitro manipulation, subsequent peptide purification, as a reporter activity for diagnostics, cell entry, etc.

5 Polypeptides selected by SPERT can be produced by any peptide synthetic method desired. Chemical synthesis can be accomplished since the amino acid sequence of the polypeptide is readily obtainable from the nucleotide sequence of the coding mRNA. Since cDNA from the coding mRNA is 10 available, the polypeptide can also be made by expressing the cDNA in a suitable host cell.

15 It is an important and unexpected aspect of the present invention that the methods described herein can be employed to identify, isolate or produce polypeptide molecules which will bind specifically to any desired target molecule. Thus, the present methods can be 20 employed to produce polypeptides specific for binding to a particular target.

Proteins contain within their primary sequence the information required to form an extraordinary variety of three dimensional shapes 25 as is well known in the art. From this variety of potential shapes, along with the charge and/or hydrophobic qualities of amino acids, comes the potential for protein functions that are used in the biosphere. Proteins provide catalysis when embodied as enzymes; proteins can provide stable 30 biological structures, for example, when used to construct spores, membranes, or viruses; and proteins can provide binding to a variety of targets, with appropriate affinities and kinetic

-44-

parameters to allow life.

Nevertheless, this vast potential in chemical activities, including the extreme potential inherent in the mammalian immune system, has actually been explored rather narrowly by organisms. This fact can be noted with a simple calculation. If the average length of a protein is 300 amino acids, and if there are twenty natural amino acids used to construct modern proteins, the number of possible sequences of proteins of average size is  $20^{300}$  or  $\sim 10^{400}$ . Estimates of the number of particles in the universe are in the range  $10^{80}$ , while estimates for the number of proteins ever explored in the entire history of the earth are in the range  $10^{10}$ . The tiny fraction of so-called sequence space that has been explored by biology is a result of evolutionary history and the relatively short age of the earth. The present invention provides the means to explore protein sequence space without historical and evolutionary limitations, while continuing to respect limitations established by the number of particles in the universe. The invention provides the means to identify and isolate polypeptide ligands with any desired quality from vast mixtures of protein sequences comprised largely of individual entities that have never before existed. The amino acid sequence of the selected ligand can be learned from the nucleotide sequence of its encoding mRNA, making tedious amino acid sequence determination unnecessary.

Even where the binding functions selected by SPERT have known naturally occurring

-45-

counterparts, there is no reason to expect that the polypeptides selected by SPERT will resemble naturally-occurring proteins or peptides having similar function. In most instances, SPERT-selected polypeptides will be smaller than naturally-occurring proteins typically having a size of from 4-100 amino acids, preferably from 4-50 amino acids selected from randomized sequence of the same length, and also having a C-terminal trailer of about 30-40 amino acids and, optionally a N-terminal leader of about 10 amino acids, for a total length of about 100 amino acids, corresponding to a molecular weight of about 11kd. This is smaller than most enzymes and all antibodies, for comparison, IgG has a molecular weight of about 150kd. Furthermore, many polypeptide ligands of the invention will function when freed by N- and C-terminal trailers. Therefore, the final product can be as small as 4-50 amino acids. The polypeptides of the invention are non-naturally-occurring, and typically differ in amino acid sequence and molecular size from naturally-occurring proteins. That portion of the amino acid sequence arising from randomized coding is designated the "binding segment" herein. The binding segment can be of any length, conveniently ranging from about 4-100 amino acids in length, preferably from about 15-50 amino acids in length. Additionally, given the vastness of sequence space, it is expected that most polypeptide ligands of the invention will have less than 50% homology with natural proteins, and preferably less than 30% amino acid homology with natural proteins.

-46-

A polypeptide ligand of the invention in a number of ways functionally resembles an antibody. Polypeptide ligands which have binding functions similar to those of antibodies can be isolated by the methods of the present invention. Such polypeptides are generally useful in applications in which polyclonal or monoclonal antibodies have found application. However, the polypeptide ligands of the invention have significant advantages over antibodies: they can be selected for any desired affinity, including higher affinities than are obtainable with antibodies, they can be selected to bind at any desired epitope or combination of epitopes, including binding sites not recognized by antibodies, they can be larger or smaller and have different solubility properties than antibodies and they can be generated by techniques that operate entirely *in vitro*, without the need for live animals or cell culture techniques.

Applications of polypeptide ligands include the specific, qualitative or quantitative detection of target molecules from any source; purification of target molecules based on their specific binding to the polypeptide; and various therapeutic methods which rely on the specific direction of a toxin or other therapeutic agent to a specific target site. Target molecules are preferably proteins, but can also include among others carbohydrates, nucleic acids, peptidoglycans and a variety of small molecules. As with conventional antibodies, polypeptide ligands can be employed to target biological structures, such as cell surfaces or viruses, through specific interaction

-47-

with a molecule that is an integral part of that biological structure. Polypeptide ligands are advantageous in that they are not limited by self tolerance, as are conventional antibodies. Also, 5 as noted, polypeptide ligands of the invention do not require animals or cell cultures for synthesis or production, since SPERT is a wholly in vitro process. The methods of the present invention related to the use of polypeptide ligands can 10 generate novel polypeptides that bind targets for which other proteinaceous ligands are known. For example, a number of proteins are known to function via binding to nucleic acid sequences, such as regulatory proteins which bind to nucleic 15 acid operator sequences. The known ability of certain nucleic acid binding proteins to bind to their natural sites, for example, has been employed in the detection, quantitation, isolation and purification of such proteins. The methods of 20 the present invention related to the use of polypeptide ligands can be used to make novel nucleic acid binding ligands having affinity for nucleic acid sequences which are known to bind proteins and to nucleic acid sequences not known 25 to bind proteins. Novel, non-naturally-occurring polypeptides which bind to the same binding sites of nucleic acids can be developed using SPERT. As will be discussed below, certain polypeptides isolatable by SPERT can also be employed to affect 30 the function, (for example inhibit, enhance or activate) specific target molecules or structures. Specifically, polypeptide ligands can be employed to inhibit, enhance or activate the function of proteins and of nucleic acids.

-48-

It is a second important aspect of the present invention that the methods described herein can be employed to identify, isolate or produce polypeptide molecules which will bind specifically to a particular target molecule and affect the function of that molecule. In this aspect, the target molecules are again preferably proteins or nucleic acids, but can also include, among others, carbohydrates and various small molecules to which specific polypeptide binding can be achieved. Polypeptide ligands that bind to small molecules can affect their function by sequestering them or by preventing them from interacting with their natural ligands. For example, the activity of an enzyme can be affected by a polypeptide ligand that binds the enzyme's substrate. Polypeptide ligands of small molecules are particularly useful as reagents for diagnostic tests, or other quantitative assays. For example, the presence of controlled substances, bound metabolites or abnormal quantities of normal metabolites can be detected and measured using polypeptide ligands of the invention. Antibodies to polypeptide ligands can be used to precipitate or bind ligand-target pairs to a solid phase matrix in a diagnostic assay. A polypeptide ligand having catalytic activity can affect the function of a small molecule by catalyzing a chemical change in the target. The range of possible catalytic activities is at least as broad as that displayed by natural proteins.

The strategy of selecting a ligand for a transition state analog of a desired reaction is one method by which catalytic polypeptide ligands

-49-

can be selected. Polypeptide ligands with high affinity for transition-state analogues are likely to have enzymatic activity, as has been demonstrated for monoclonal antibodies directed  
5 against transition-state analogues. These antibodies have exhibited a wide range of catalytic activities, including acyl-transfer reactions [Pollack et al., Science 234:1570 (1986); Tramantano et al., Science 234:1570  
10 (1986); Jacobs et al., J. Am. Chem. Soc. 109:2174 (1987); Napper et al., Science 237:1041 (1987); Janda et al., Science 241:1188 (1988); Schultz,  
P.G., Science 240:426 (1988); Benkovic et al., Proc. Natl. Acad. Sci. 85:5355 (1988)], carbon-  
15 carbon bond formation [Jackson et al., J. Am. Chem. Soc. 110:4841 (1988); Hilvert and Nared, J. Am. Chem. Soc. 110:5593 (1988)], carbon-carbon bond cleaving reactions [Cochran et al., J. Am. Chem. Soc. 110:7888 (1988)], peptide cleavage  
20 [Iverson and Lerner, Science 243:1184 (1989)], and ester bond hydrolysis [Janda et al., Science 244:437 (1989)]. The number of polypeptide sequences and structures that can be explored by SPERT far exceed those available in the immune  
25 system.

Enzymes are evolved using SPERT and starting randomized sequences corresponding to about 50 amino acids, as in Example 3. Enzymatic polypeptide ligands of small size are entirely  
30 unanticipated by the present understanding of enzymology; enzymes are always much larger in nature than the scientist expects. The specific transition state analogues used are drawn from the literature cited above. Among the reactions

-50-

probed by the monoclonal antibody-enzymes are some which lead to the breakdown of toxic waste products, including chemicals with chlorine-carbon bonds and carbon-carbon bonds in ring structures like those found in benzene and polychlorinated phenols.

The binding selection methods of the present invention can be combined with secondary selection or screening to identify ligands capable of modifying target molecule function upon binding. The large population of variant amino acid sequences that can be tested by SPERT enhances the probability that polypeptide sequences can be found that have a desired binding capability and that function to modify target molecule activity. The methods of the present invention are useful for selecting polypeptide ligands which can selectively affect function of any target protein. The methods described herein can be employed to isolate or produce polypeptide ligands which bind to and modify the function of any protein or nucleic acid. It is contemplated that the method of the present invention can be employed to identify, isolate or produce polypeptide molecules which will affect catalytic activity of target enzymes, i.e., inhibit catalysis or modify substrate binding, affect the functionality of protein receptors, i.e., inhibit binding to receptors or modify the specificity of binding to receptors; affect the formation of protein multimers, i.e., disrupt quaternary structure of protein subunits; and modify transport properties of protein, i.e., disrupt transport of small molecules or ions by proteins.

-51-

Secondary selection methods that can be combined with SPERT include among others selections or screens for enzyme inhibition, alteration of substrate binding, loss of functionality, disruption of structure, etc.

5 Those of ordinary skill in the art are able to select among various alternatives those selection or screening methods that are compatible with the methods described herein.

10 An embodiment of the present invention, which is particularly useful for identifying or isolating polypeptides which bind to a particular functional or active site in a protein, or other target molecule, employs a molecule known, or selected, for binding to a desired site within the target protein to direct the selection/amplification process to a subset of polypeptide ligands that bind at or near the desired site within the target molecule. In a simple example, a polypeptide sequence known to bind to a desired site in a target molecule is incorporated near the randomized region of all polypeptides being tested for binding. SPERT is then used to select those variants, all of which will contain the known binding sequence, which bind most strongly to the target molecule. A longer binding sequence, which is anticipated to either bind more strongly to the target molecule or more specifically to the target can thus be selected. The longer binding sequence can then be introduced near the randomized region of the polypeptide test mixture and the selection/amplification steps repeated to select an even longer binding sequence. Iteration of

15

20

25

30

-52-

these steps (i.e., incorporation of selected sequence into test mixtures followed by selection/amplification for improved or more specific binding) can be repeated until a desired level of binding strength or specificity is achieved. This iterative "walking" procedure allows the selection of polypeptides highly specific for a particular target molecule or site within a target molecule. Another embodiment of such an iterative "walking" procedure, employs an "anchor" molecule which is not necessarily a polypeptide or amino acid. In this embodiment a molecule which binds to a desired target, for example a substrate or inhibitor of a target enzyme, is chemically modified such that it can be covalently linked to a bridge molecule which in turn is known to be bound to an oligopeptide of known sequence. The bridge molecule covalently linked to the "anchor" molecule that binds to the target also binds to the target molecule. The sequence encoding the known bridge-binding oligopeptide is incorporated near the randomized region of the test nucleic acid mixture. SPERT is then performed to select for those polypeptide sequences that bind most strongly to the target molecule/bridge/anchor complex. The iterative walking procedure can then be employed to select or produce longer and longer polypeptide molecules with enhanced strength of binding or specificity of binding to the target. The use of the "anchor" procedure is expected to allow more rapid isolation of polypeptide ligands that bind at or near a desired site within a target molecule. In particular, it is expected that the "anchor"

-53-

method in combination with iterative "walking" procedures will result in polypeptides which are highly specific inhibitors of protein function.

In accordance with the teachings of copending application Serial No. 07/536,428, the translated mRNA of a ribosome complex is, in principle, capable of binding to target molecules and of being partitioned concurrently with nascent polypeptides. In particular, where partitioning is accomplished by affinity chromatography, the selected ligand can be an RNA, rather than a polypeptide. Binding of mRNA can be differentiated from polypeptide binding once the ligand has been selected and both the selected polypeptide and its coding mRNA are available for independent direct binding studies where the two are not part of the same ribosome complex. Comparative studies of the relative frequency of RNA ligands and polypeptide ligands selected by SPERT are of fundamental biological importance to understanding the specialization of function that currently exists in living cells.

The polypeptides of the invention can be selected for other properties in addition to binding. For example, during partitioning, stability to certain conditions of the desired working environment of the end product can be included as a selection criterion. If a polypeptide which is stable in the presence of a certain protease is desired, that protease can be part of the buffer medium used during partitioning. As will be understood, conditions which disrupt ribosome complexes should be avoided. Other desired properties can be

incorporated, directly into the polypeptide sequence as will be understood by those skilled in the art. For example, membrane affinity can be included as a property, either by employing a N- or C-terminal trailer having high hydrophobicity, or by biasing the randomized coding to favor the amino acids with lipophilic side chains.

The coding nucleic acid concomitantly selected by partitioning nascent polypeptides as described, is useful in its own right to transform host cells or organisms. The transformed organism is then useful for, e.g., fermentation production of the selected polypeptide. A transgenic organism can be rendered resistant to a virus infection, for example, by causing *in vivo* synthesis of a polypeptide ligand of the viral nucleic acid or a key viral protein. In principle, any functionality contributed by a polypeptide ligand of the invention can be bestowed on a suitable host organism. Methods known in the art can be used to combine the coding region with a promoter, polyadenylation signal functional in the intended host, followed by incorporation into a suitable vector for transformation, all as known and understood in the art.

Table 1

- 1.) 5'-CCGAAGCTTAATACGACTCACTATAGGGCGACATAACATTACACACATAA-3'
- 2.) 5'-CGGGAAATTCTTCTATATTATTTCCCTTATGTGTGTAAATGTATG-3'
- 3.) 5'-GGCGAATTCTGCTGCTGCAGTGCTGCCATGGTTGCGACGGTCAGGA-3'
- 4.) 5'-CCGCCGGATCCTCCTGTCCGTCGCAA-3'
- 5.) 5'-CCCGAATTTC-[-45N-]-CTGCAGTGCTGCCATGGT-3'
- 6.) 5'-ACCATGGCAGCACTG-3'
- 7.) 5'-GGGCCATGG-[-120(ACG)-]-CCATGGTTGCGATGGTCAGGA-3'
- 8.) 5'-TCCTGTCCATCGCAA-3'

-55-

TABLE 2

## 1.) 5' fixed sequence

HindIII site	+1	Ribosome binding site	EcoRI
<hr/>			
5'-CCGAAGCTTAAATACGACTCACTATAGGGCGACATACATTTACACACATAA	Aggaggaaauauauatgaaagaattcccg-3'		
3'-ggcttcgaattatgtcgactatcccgt	GTATGTAAATGTGTATTCCCTTATATTATACTTCTTAAGGGC-5'		
- T7 promoter -			

## 2.) Stratagene polylinker cloning site (pBSSK+)

PstI	<hr/>		
5'-TCGATAAGCTTGATATCGAATTCTGCAGCCCCGGGGATCCACTAG-3'			
<hr/>			
HindIII	EcoRI	BamHI	

## 3.) 3' primer annealing site and insertion sequence cloning sites

EcoRI	PstI	NcoI	BamHI
<hr/>			
5'-GGCGAATTCTGCTGCTGCAGTGCTGCCATGGTTGCAGCGTCAGGA	ggatccggcgg-3'		
3'-ccgcctaagacgacgacgtcacgacggtaacc	AACGCTGCCTGTCCTCCTAGGCCGCC-5'		

## 4.) Randomizing oligonucleotides to be cloned at the EcoRI, PstI, and NcoI sites.

EcoRI	PstI	
<hr/>		
a.) 5'-CCCGAATT-[ <del>-45N-</del> ]-CTGCAGTGCTGCCATGGT-3'		
3'-GTCACCGACGGTACCA-5'		

NcoI	NcoI	
<hr/>		
b.) 5'-GGGCCATGG-[ <del>-120(ACG)-</del> ]-CCATGGTTGCATGGTCAGGA-3'		
3'-AACGCTACCTGTCCT-5'		

-56-

EXAMPLES

The techniques and methods used in the ensuing examples are published and known in the art. Together with adaptations and modifications known to those of ordinary skill in the art, the procedures not specifically referenced herein are available from known reference works. In addition to Sambrook et al., (1989) supra, Genetic Engineering, Plenum Press, New York (1979); Weir, (ed.) (1986) Handbook of Experimental Immunology in Four Volumes, 4th Ed, Blackwell Scientific Publications, Oxford; and the multivolume Methods in Enzymology published by Academic Press, New York. Polymerase chain reaction techniques are described in PCR Protocols (Michael A. Innis, et al. eds.) (1990) Academic Press, Inc.

Throughout the following examples, reference is made to Tables 1 and 2. Table 1 lists oligonucleotide sequences used for preparing mRNA candidates. Table 2 lists the same sequences together with explanatory notes showing functional domains. Sequences in capitals are chemically synthesized, sequences in lower case letters are complementary sequences made enzymatically by DNA polymerase.

Example 1.      Direct Immunoprecipitation of  
                    Ribosome Complexes: Polypeptide  
                    Ligands Directed Toward  
                    Immunoglobulin Molecules.

The method of the invention is used to select novel polypeptides that bind the antibody of an epitope commonly recognized by the antisera

-57-

from autoimmune mice which are the f1 progeny of a cross of NZB and NZW parents (Portanova et al., J. Immunol. 144, 4633 (1990)). The known epitope consists of about 10 contiguous amino acids at the 5 amino terminus of the histone H2B protein. To make mRNA encoding candidate polypeptides, a 5' fixed sequence composed of a T7 promoter sequence and a ribosome binding site which is recognized by both prokaryotic and eukaryotic ribosomes, 10 terminating in a restriction endonuclease site is synthesized and cloned using oligonucleotides having the sequences shown as sequence 1 in Tables 1 and 2 and in Figure 8. A 3' fixed sequence is placed into a restriction site to provide an mRNA 15 encoding the C-terminal trailer sequence of ca. 100 nucleotides lacking stop codons (for ca. 30-35 amino acids) shown as sequence 3 in Tables 1 and 2 and Figure 8. In addition, as shown in Figure 1, a 3' primer annealing site (sequence 3) is 20 provided so that cDNA synthesis can be accomplished on the mRNA recovered from partitioned ribosome complexes.

The randomized polypeptide insertion site is bounded by restriction endonuclease 25 recognition sites, in this example EcoRI and PstI. A single-stranded oligonucleotide is synthesized with a randomized sequence of 45 nucleotides 30 (corresponding to 15 codons) bounded by specific sequences that include those two restriction endonuclease sites (Sequence 4a). Synthesis of randomized oligonucleotides is carried out using an Applied Biosystems DNA synthesizer provided with a reactant mixture for each nucleotide position. To partially compensate for the amino

-58-

acid sequence bias inherent in the redundancy of the genetic code, the reaction mixtures contain, on a mole percent basis, the following composition of bases for each codon: First position, C-20%, T, A, and G-30% each; Second position, C-15%, A-35%, T and G-25% each; Third position, T, C, A and G-25% each. Using a nucleic acid primer that is complementary to the fixed 3' end of the randomized oligonucleotide, randomized double-stranded DNA is created with the action of DNA polymerase. The products are digested with the two restriction endonucleases and ligated between the 5' fixed sequence and the 3' fixed sequence discussed above. In vitro transcription of these ligated templates using T7 RNA polymerase (Bethesda Research Laboratories, Gaithersburg, MD) provides mRNA templates for in vitro translation. A rabbit reticulocyte lysate system (BRL) is used to translate the mRNA templates in vitro, using standard reaction conditions. Such translation of these transcripts results in a variety of ribosomal complexes (mRNA-nascent polypeptide-tRNA-ribosomes) that are identical except for the randomized region of the nascent polypeptide.

Antibodies (IgGs), Portamova et al., supra, which recognize the H2B histone epitope are added to the in vitro translation mixture. Immunoprecipitation of the immunoreactive ribosome complexes partitions the mRNAs species that encode the highest-affinity polypeptide ligands in the population (see Figures 3 and 4). Immunoprecipitated complexes are separated by low speed centrifugation. cDNA is synthesized from these mRNAs and is used via PCR to provide

-59-

template for further cycles of transcription,  
translation, immunoselection and cDNA synthesis.

Clones are isolated as described in  
Application 07/536,428, June 11, 1990,  
5 incorporated herein by reference. The individual  
polypeptide products are over produced and  
purified and tested, using standard techniques for  
reactivity to the anti-H2B histone antibodies. In  
addition, the polypeptide ligands are challenged  
10 competitively with authentic histone H2B-derived  
epitopes to discover which polypeptide ligands  
bind to the same portion of the antibodies as the  
true epitope. Among the polypeptides isolated  
that bind the antibody are found those having less  
15 than 50% sequence homology with the H2B histone  
epitope. Other antibody binding sequences are  
identified having less than 30% homology with the  
H2B histone epitope. Other polypeptide ligands of  
the antibody do not compete for the H2B epitope  
20 binding site.

Example 2. Diagnostics using the polypeptide  
ligands of Example 1: An assay for  
25 anti-H2B antibodies in the progeny  
of NZB X NZW mice.

Auto-immune diseases result from the  
elaboration of an inappropriate antibody molecule  
with reactivity toward a normal cellular component  
30 (often a protein, but sometimes a nucleic acid, as  
in Systemic Lupus Erythematosis - SLE).  
Polypeptide ligands generated through the SPERT  
protocols in Example 1 are aimed at diagnosis of  
mouse "Lupus" in the offspring of NZB X NZW mice.

-60-

SPERT is used to identify and obtain a reagent ligand for the diagnostic recognition of the auto-antibody that recognizes the histone H2B epitope.

As in Example 1, ribosome complexes are treated with the auto-antibody to partition reactive polypeptides from non-reactive polypeptides resident (as nascent polypeptides) in ribosome complexes. The auto-antibodies are used to precipitate the ribosome complexes containing polypeptides that fit into the active site of the antibody. The most avidly bound polypeptide emerges from repeated SPERT cycles.

The most avidly bound polypeptide ligand does not resemble in detail the epitope identified as the portion of the target that reacts with the antibody. Auto-immune diseases are triggered by unknown antigens, which are not necessarily the same as the target/epitope identified as the interactive species during the clinical stage of the auto-immune disease. For example, a virus infection may trigger an immune reaction that yields a class of antibodies that cross-react with a normal cellular target. Such antibodies may bind more avidly to the original, stimulatory, viral antigen than to the epitope on the cellular target. As another example, the epitope on the cellular target may not take full advantage of the binding site on the antibody.

The polypeptide ligand is used diagnostically to measure the quantity of circulating auto-antibody, using, e.g., an ELISA assay. The technology is available to one skilled in the art, without undue experimentation. As another example, the fixed portion of the

-61-

polypeptide ligand is used as the reporter substance when the polypeptide ligand interacts with the circulating auto-antibody. With a fixed carboxy-terminus of beta-galactosidase or alkaline phosphatase, serum protein samples attached to plastic plates are assayed directly for the anti-H2B antibody by "staining" with the polypeptide ligand covalently fused (by recombinant DNA techniques) to either reporter enzyme.

10

Example 3. Indirect Immunoprecipitation:  
Polypeptide ligands directed toward domains of any protein.

15

Immunization of animals with antigens, whether crudely prepared or purified, often results in immune responses directed at a subset of the available epitopes in that antigen. The polyclonal sera may react largely with a single protein domain in that antigen. Similarly, when researchers attempt to raise antibodies against fusion proteins, often the well-known fusion partner is immuno-dominant over the new protein portion of the fusion.

20

Antibodies aimed at a protein target (but that do not recognize the portion of the target that one wishes to use as the target in SPERT) allow INDIRECT Immunoprecipitation of ribosome complexes. That is, immunoprecipitation is a useful partitioning step when antibodies are aimed at domains in the target that are different from those domains pre-selected for SPERT-based ligand evolution. This protocol is sometimes called "panhandling", and can yield high-affinity

30

-62-

polypeptide ligands for target domains that are weakly immunogenic.

SPERT is performed using variable material prepared as in Example 1 except that the randomized mRNA regions are now set to yield about 50 amino acids in the solvent-exposed nascent polypeptide. Biased randomization is done so that chain termination codons are not likely over the 150 randomized nucleotides; in addition, cell-free translation is performed in the presence of so-called suppressor tRNAs so that translation continues to the desired portion of the mRNAs.

The population of ribosome complexes is pre-treated with the antisera aimed at the target protein, but in the absence of that target protein. The pre-treatment is designed to eliminate any nascent polypeptides that react directly with the antibodies, as in Example 1. The target protein is then added to the ribosome complexes, along with antibodies aimed at the target protein. Partitioning occurs as the ribosome complexes that interact with the target at the same time (see Figure 4).

The single-stranded DNA binding protein of bacteriophage T4 (gp32) has an acidic carboxyterminal region which is immunodominant (K. Krassa, Ph.D., Thesis, 1987). In one immunization experiment, polyclonal sera react exclusively with the carboxyterminal domain of the protein; 12 monoclonal cell lines derived from hybridoma fusions with 12 monoclonal cell lines derived from hybridoma fusions with spleen cells from such immunized animals produced antibodies that react with the same target domain. Purified polyclonal

-63-

sera which react with the carboxy-terminal domain of gp32 are used for indirect immunoprecipitation in this example.

A population of ribosome complexes is produced (above). These ribosome complexes are pre-treated with the polyclonal sera aimed at gp32; this is readily accomplished by passing the ribosome complexes through Staph A columns pre-bound with the polyclonal sera against gp32. Subsequently, those ribosome complexes unable to react directly with antibodies raised against gp32 are reacted with gp32, followed by treatment with the sera aimed at the carboxy-terminus of gp32. Goat anti-mouse antibodies are used to immunoprecipitate gp32 and whatever ribosomal complexes interact with the core domain of gp32. Cycles of SPERT are continued until a desired level of binding is attained. Sequences are then cloned and individuals identified and tested for affinity to gp32.

Example 4. Isolation of a polypeptide ligand for a serine protease.

Serine proteases are protein enzymes that catalyze hydrolysis of peptide bonds within proteins, often with high selectivity for specific protein targets (and, of course, for specific peptide bonds within the target protein). The serine proteases are members of a gene family in mammals. Examples of serine proteases are tissue plasminogen activator, trypsin, elastase, chymotrypsin, thrombin, and plasmin. Many disease states can be treated with polypeptide ligands

-64-

that bind to serine proteases, for example, disorders of blood clotting. Elastase inhibitors are likely to be useful in minimizing the clinical progression of emphysema. Proteases other than 5 serine proteases are also important in mammalian biology, and these too are targets for polypeptide ligands with appropriate affinities obtained according to the invention herein taught.

A ligand that binds to porcine elastase 10 is identified and purified using the starting randomized material of Example 3. Serine proteases are easily attached by standard methods to column support materials with retention of enzymatic activity. Porcine elastase attached to 15 agarose is available from commercial sources. Thus, in this example affinity chromatography is the partitioning method. Natural elastase inhibitors are available, and are used to check that the active site of the bound elastase is 20 available for the binding of an inhibitory ligand. The buffer used for binding during the SPERT cycles must not denature or otherwise inactivate elastase; dithiothreitol, which can reduce protein disulfide bonds, is left out of the binding 25 buffer.

After several rounds of SPERT, as the affinity of the mixture of nascent polypeptides becomes high, a reversal of the elution parameters is used. Early rounds of SPERT are aimed at 30 obtaining any polypeptide ligand that binds to any domain of elastase; after virtually all the nascent polypeptides are able to bind the column, the ribosome complexes are poured through a column that has been pre-saturated with a natural

-65-

inhibitory ligand for the elastase active site. In addition, the elution buffer for this procedure includes high concentrations of that same natural inhibitory ligand. The ribosome complexes that  
5 are not bound in this reversed elution procedure are used to prepare mRNAs for further SPERT cycles, once again depending on high affinity for the bound elastase. This procedure focuses the evolving polypeptide ligands toward the elastase  
10 active site.

When the mixture of polypeptide ligands has a high affinity for the bound elastase, and is aimed primarily toward the active site, further enrichment for high affinity inhibitors of  
15 elastase activity is accomplished by including low concentrations of the natural inhibitors in the partitioning steps, thus demanding that the evolving polypeptide ligands have higher affinity than the effective affinity of the natural  
20 inhibitor at the concentration used.

Nucleic acids encoding polypeptide ligands are cloned and sequenced, and binding affinities and inhibitory binding affinities for elastase are measured. Binding affinities and  
25 inhibitory efficiencies are measured with the same polypeptide ligands for other members of the serine protease family in order to ascertain specificity within the family.

30 Example 5. Polypeptide ligands that antagonize a receptor: A synthetic inhibitor of the interleukine 1 receptor.

Receptors are a class of proteins that

-66-

are partially integrated into the cell's cytoplasmic membrane such that a domain resides outside the cell. That domain serves as a binding site for cell extrinsic molecules, including growth factors, peptide hormones, non-peptide organic molecules (which may include hormones), or even ions. Receptors handle the bound ligand in several different ways, including signal transduction through the membrane or internalization of the bound ligand for its subsequent function. In either case polypeptide ligands of the invention may be used to affect function of the receptor, that is to cause the normal activity of the natural ligand or to block that activity.

Receptor antagonism for a useful therapeutic purpose is accomplished by generating a polypeptide ligand through SPERT that is aimed at the interleukine 1 (IL-1) receptor. A natural antagonist of the receptor has been found (Hannum et al., *Nature*, 343:336-340 (1990); Eisenberg et al., *Nature*, 343:341-346 (1990), and that antagonist has the presumptive utility of preventing or easing inflammatory problems such as those found in rheumatoid arthritis. The natural antagonist (called IL-1ra for IL-1 receptor antagonist) is partially homologous to IL-1 itself, and is a competitive inhibitor of interleukine 1 binding to the receptor. The natural IL-1ra is a pure antagonist, completely without agonist activity at the highest concentrations used in the work cited above. IL-1ra is synthesized as a protein with 177 amino acids; after post-translational cleavage the

-67-

active inhibitor has 152 amino acids and,  
additionally, is glycosylated. However, the  
activity of recombinant IL-1ra, without  
glycosylation, is comparable to the activity of  
the natural inhibitor.

SPERT is used to develop a polypeptide  
ligand antagonist for the interleukine 1 receptor.  
Two methods are used. In the first monoclonal  
antibodies are raised against interleukine 1 that  
are able to cross-react with IL-1ra. Such  
monoclonal antibodies in principle recognize the  
features in common between IL-1 and IL-1ra. Those  
monoclonal antibodies are used, as in Example 1,  
to develop polypeptide ligands that bind to the  
antigen combining site; such polypeptide ligands  
are candidates for a novel class of IL-1  
antagonists. Since one goal in this case is to  
provide antagonists smaller than the natural IL-  
1ra, the randomized polypeptide is ca. 50 amino  
acids, as in Example 3.

In a second methodology the  
extracellular domain of the IL-1 receptor is  
itself used as the target for polypeptide ligand  
development through SPERT. The domain is attached  
to an insoluble matrix. Candidate polypeptide  
ligands, residing in ribosome complexes, are  
partitioned on the matrix. The matrix is eluted  
with high concentrations of IL-1, thus displacing  
the ribosome complexes and nascent polypeptides  
with the natural ligand known to bind to the  
desired active site on the receptor. Cycles of  
SPERT are continued until high affinity  
polypeptide ligands are identified.

Very high affinity, even covalent,

-68-

antagonists of the receptor are isolated by an elution protocol during SPERT that denatures the ribosome complexes even if the polypeptide ligand remains strongly bound to the receptor. The mRNA eluted from the column under protein denaturing conditions is used to prepare cDNA which is amplified through PCR, after which transcription provides mRNA for the next round of SPERT.

All genes encoding polypeptide ligands are sequenced, and the polypeptide ligands are tested for IL-1 receptor antagonism. Those ligands identified by receptor-based affinity chromatography are tested with the antibodies of the first method to screen for the novel antagonists recognized by those antibodies that recognize structural or sequence homology between IL-1 and IL-1ra. Novel, SPERT-generated polypeptide ligands having IL-1 receptor antagonist activity are isolated and characterized. SPERT-generated antagonists having less than 50% amino acid homology with natural IL-1ra are identified. In addition, SPERT-generated antagonists having less than 30% amino acid homology are identified.

Example 6. Protein improvement by SPERT:  
Mutagenesis and selection of better natural insecticides.

Bacillus thuriengiensis is a gram-positive, spore-forming bacteria which produces insecticidal proteins. These proteins, derived from different B. thuringiensis strains, have varying effectiveness for killing insect larvae of

-69-

different species. Although one specific protein will kill the insect larvae of a variety of species, the effectiveness toward the different insect targets (measured as the level of protein required to produce 50% mortality) can vary by as much as 2000-fold. The mechanism of action for these insecticide proteins is to bind a receptor on the gut membranes of the susceptible insect larva. Such membranes serve as a functional partitioning tool in SPERT.

We create double-stranded DNA templates suitable for SPERT by PCR; the appropriate DNA encodes the N-terminal 646 amino acid portion of the insecticidal protein from *t. subspecies*

*kurstaki* HD-1, which is fully active (Fischhoff et al., Biotechnology 5:807-813 (1987)). This protein kills the larva of tomato hornworm and cabbage looper very effectively at low concentration. Substantially more protein is required to kill tobacco budworm, corn earworm, black cutworm, European cornborer, and beet armyworm. Gut membranes from each of these insect larvae will be used as partitioning agents in SPERT.

The starting material in these experiments is RNA derived from the cloned gene, as above. Two methods are used to create protein variants. In one method mutagenic PCR provides random mutations throughout the 646 amino acids of the insecticide. In fixed codons within the insecticide, using about 50 amino acid replacements. In particular, randomized DNA is used to replace the codons encoding the hypervariable region of the *Bt.* toxin. Rounds of SPERT are continued until a desired level of

-70-

binding to gut membranes is achieved. The DNA products are cloned and sequenced and individually assayed for effectiveness in binding membranes and larval killing. Effective toxins are selected by  
5 SPERT, having a naturally-occurring sequence replaced by a sequence that is less than 50% homologous with the replaced sequence. In addition, toxic, SPERT-generated variants are identified wherein the original, naturally-  
10 occurring sequence is replaced by a sequence having less than 30% sequence homology with the replaced sequence.

Example 7. Anti-viral polypeptide ligands:  
15 Inhibition of viral entry into target cells.

Receptors are often used for viral attach on cells. Recently Kaner et al. (Science, 20 248:1410-1413 (1990)) described the basic fibroblast growth factor (FGF) receptor as the likely portal through which Herpes Simplex Virus Type 1 (HSV) enters a cell. In that same paper, by citation of other work several other viruses 25 are said to utilize other receptors to gain cellular entry. Rhinovirus, the common cold virus, is said to enter cells through a cell adhesion molecule ICAM-1. HIV, the AIDS virus, enters cells through the CD4 glycoprotein receptor. Epstein-Barr virus enters T lymphocytes via the C3d complement receptor. Rabies virus enters nerve cells through the acetylcholine receptor. Reovirus enters cells through the beta-adrenergic receptor. Vaccinia virus enters cells  
30

-71-

through a functional interaction with the epidermal growth factor receptor. Apparently viruses survive in part by using absolutely crucial cell receptors to gain entry into susceptible hosts. That is, host organisms can not easily alter such important receptors so as to become resistant to the virus without suffering some impairment of crucial cell and organism functions.

Polypeptide ligands of the invention are identified that diminish viral uptake through receptors while still allowing critical growth factors to function. The basic FGF receptor is used to demonstrate a successful strategy. The soluble domain of the basic FGF receptor (Lee et al., *Science*, 245:57 (1989)) is used as the target. A candidate mixture of polypeptide ligands is used as in Example 3. The partitioning of ribosome complexes is obtained with matrix bound extracellular domain of the FGF receptor. The cycles of SPERT are altered to include an elution step from the matrix with high concentrations of HSV; during this elution step the ribosome complexes that exit the column are discarded, while those ribosome complexes that remain on the column are further eluted with high concentrations of FGF itself. Those ribosome complexes that are not displaced by HSV but are displaced by FGF contain nascent polypeptides that are candidates ligands with the desired specificity. Such polypeptides bind FGF receptors in a way that inhibits HSV binding but does not interfere with FGF binding. Several cycles of SPERT are used to find the most avidly bound

-72-

polypeptide that is eluted with FGF but not with HSV. Candidate polypeptides are assayed for their negative impact on HSV infection and their inability to prevent FGF-mediated cell growth.

5       The most useful polypeptide ligands in this example are neither antagonists nor agonists of the FGF receptor at concentrations that diminish HSV infection. Novel polypeptides meeting these criteria are made using the process as described.

10      A polypeptide meeting the criteria having less than 50% amino acid homology with FGF is isolated. In addition, a polypeptide meeting the criteria having less than 30% homology with FGF is isolated.

15      Example 8.     Polypeptide ligands that enter cells: The glucocorticoid receptor and trojan horse ligands.

20      The glucocorticoid receptor protein binds steroid hormone, after which the receptor protein is internalized from the membrane so that the receptor can make its way into the cell nucleus. The receptor has a DNA binding domain (DBD) that interacts in the nucleus with target DNA sequences. Polypeptide ligands of the invention, agonists of the glucocorticoid receptor, are internalized along with the receptor, and thus directed sequentially to the cytoplasm and then to the nucleus. Depending on the dissociation rate constant for specific polypeptide ligands, these ligands largely reside after uptake in either the cytoplasm or the nucleus.

-73-

Using the randomized starting material of Example 3, SPERT is directed toward the glucocorticoid receptor, either with indirect immunoprecipitation or affinity chromatography using bound receptor. As in prior example, SPERT protocols are manipulated so that polypeptides are found that compete directly for the glucocorticoid binding domain but that have much lower affinity than that observed for steroid hormones. As the polypeptide ligands evolve, screening of potential ligands is performed on individual candidates; thus resistance to proteolysis of the polypeptide ligand is tested using whole cell entry prior to the protease challenge, and testing both cells with and without an abundance of the glucocorticoid receptor. Polypeptide ligands that enter cells are localized in the cytoplasm or nucleus by means available to those skilled in the art. Those polypeptide ligands that enter cells with proper localization are fused to other polypeptide ligands to provide cell entry for molecules with other useful activities.

Example 9. Polypeptide ligands toward nucleic acids: Inhibitors of transcription.

Cancer cells can result from the over-expression of a transcriptional activator protein that functions to enhance transcription and subsequent expression of sets of genes that push the cell toward inappropriate and uncontrolled growth. Thus, mutations that elevate the activity of a transcriptional enhancer may cause cancer

-74-

through enhancement of the expression of a set of genes relevant for growth control. Such tumors are treatable with polypeptide ligands that reset the appropriate level of expression or activity of the transcriptional enhancer. While it is likely that polypeptide ligands may be aimed at the enhancer protein directly, thus inhibiting the activity and resetting a proper growth rate, in the present example a polypeptide ligand is aimed 5 at the production rate of the transcriptional enhancer.

10

The polypeptide ligand of interest binds to the genome of the cancer cell at a location that competes for transcription of the gene that encodes the transcriptional activator protein, and hence expression of that protein. That is, in 15 classical genetic language, the polypeptide ligand is a specific transcriptional repressor.

The starting materials of Example 3 are used to generate a mixed pool of candidate polypeptides. A specific sequence of double-stranded DNA is prepared by chemical means and covalently attached to an insoluble column matrix. The column matrix is chosen such that ribosome complexes in general are able to flow through the column containing bound DNA. Ribosome complexes containing nascent polypeptide ligands that interact with double-stranded DNA (either with sequence specificity or not) are retarded on the 20 column, recovered, and placed into the SPERT protocol of mRNA-amplification, transcription, and a second cycle. In order to eliminate polypeptide ligands with affinity for all double-stranded DNA 25 (that is, without adequate sequence specificity

30

-75-

for the intended use), the ribosome complexes are mixed with random soluble double-stranded DNA sequences prior to the column partitioning step. The soluble DNA concentration is adjusted to give about tenfold more non-specific DNA during the partitioning step than is the abundance of specific DNA sequences attached to the column. In this manner polypeptide ligands that are indifferent to DNA sequence emerge from the column along with ribosome complexes containing polypeptide ligands that are unable to bind DNA at all.

Polypeptide ligands aimed at a specific DNA sequence are characterized further.

Randomized DNA sequences are used to establish which nucleotide pairs in the covalently attached DNA are required for avid binding of the polypeptide (using the SELEX protocol described in U.S. Patent Serial No. 07/536,428). A second SPERT is directed toward the contiguous DNA base pairs that are not bound by the first isolated polypeptide ligand, and the genes for the first and second polypeptide ligands are combined to yield a polypeptide ligand fusion (in either order, and containing a flexible peptide linker) to provide a polypeptide ligand with higher specificity and avidity than is available from either polypeptide ligand by itself. This improvement in specificity and avidity is an example of walking, although in this case the "steps" are made independently and the polypeptide ligands joined post-identification.

The sequence of double-stranded DNA chosen in this example must overlap a

-76-

transcriptional initiation signal. The ras oncogene transcriptional initiation region is chosen first.

-77-

WE CLAIM:

1. A method for making a polypeptide ligand of a target molecule comprising:
  - 5 a) synthesizing a translatable mRNA mixture comprising a ribosome binding site, translation initiation codon and a randomized sequence coding region;
  - 10 b) synthesizing a mixture of ribosome complexes, each member thereof comprising a ribosome, a nascent polypeptide and a translated mRNA, said mRNA having a randomized coding region and said nascent polypeptide being the translation product of said mRNA;
  - 15 c) partitioning the ribosome complexes with respect to binding of the ribosome complexes to a desired target molecule, thereby separating the ribosome complexes into ribosome complex-target pairs and unbound complexes, the ribosome complex-target pairs having mRNA enriched for sequences encoding target-binding polypeptides;
  - 20 d) amplifying the mRNA of partitioned ribosome complex-target pairs to yield a translatable mRNA mixture comprising a ribosome binding site, an initiation codon and a coding region enriched for sequences encoding target-binding

-78-

polypeptides;

- 5           e) repeating steps b) through d) using  
the mRNA enriched for sequences  
encoding target-binding  
polypeptides of each successive  
repeat as many times as desired to  
yield a desired level of target  
binding by a polypeptide encoded by  
the mRNA enriched for sequences  
encoding the polypeptide; and  
10           f) synthesizing a polypeptide encoded  
by the enriched mRNA of step e),  
thereby making a polypeptide ligand  
of a target molecule.

- 15           2. The method of claim 1 wherein the target  
molecule is a protein.  
  
20           3. The method of claim 2 wherein the  
protein is an enzyme.  
  
25           4. The method of claim 2 wherein the  
protein is an antibody.  
  
25           5. The method of claim 2 wherein the  
protein is a receptor.  
  
30           6. The method of claim 2 wherein the  
protein is a nucleic acid binding  
protein.  
  
30           7. The method of claim 2 wherein the  
protein is a toxin.

-79-

8. The method of claim 2 wherein the protein is a glycoprotein.
9. The method of claim 2 wherein the protein is an antigen.  
5
10. The method of claim 1 wherein the polypeptide is an inhibitor of function of the target molecule.  
10
11. The method of claim 1 wherein the target molecule is a cell membrane component.  
15
12. The method of claim 1 wherein the target molecule is a virus component.  
15
13. The method of claim 1 wherein the target molecule is a carbohydrate.  
20
14. The method of claim 1 wherein the target molecule is a polysaccharide.  
25
15. The method of claim 1 wherein the target molecule is a lipid.  
16. The method of claim 1 wherein the target molecule is a glycolipid.  
30
17. The method of claim 1 wherein the target molecule is a toxin.  
18. The method of claim 1 wherein the target molecule is a drug.

-80-

19. The method of claim 1 wherein the target molecule is a controlled substance.
20. The method of claim 1 wherein the target molecule is a metabolite.
- 5 21. The method of claim 1 wherein the target molecule is a cofactor.
- 10 22. The method of claim 1 wherein the target molecule is a nucleic acid.
23. The method of claim 1 wherein the target molecule is a hormone.
- 15 24. The method of claim 1 wherein the target molecule is a receptor ligand.
- 20 25. The method of claim 1 wherein the target molecule is a transition state analog.
26. The method of claim 1 wherein the translatable mRNA mixture is synthesized by in vitro transcription of a cDNA mixture comprising additionally a transcription promoter sequence near the 5' end of the cDNA.
- 25 27. The method of claim 1 wherein the translation initiation codon is situated within 5-12 bases from the ribosome binding site in the 3' direction on the mRNA.
- 30

-81-

28. The method of claim 1 wherein the ribosome complexes are synthesized in an in vitro translation system lacking release factor.

5

29. The method of claim 1 wherein the partitioning is carried out by column chromatography.

10 30. The method of claim 1 wherein the partitioning is carried out by binding to target molecules attached to a solid phase matrix.

15 31. The method of claim 1 wherein the partitioning is carried out by immunoprecipitation.

20 32. The method of claim 1 wherein the partitioning is carried out by indirect immunoprecipitation.

25 33. The method of claim 1 wherein the mRNA is amplified in step d) by polymerase chain reaction.

30 34. The method of claim 1 wherein the process of amplifying in step d) includes introducing mutations during amplification.

35. The method of claim 1 comprising additionally, after step e) the step of cloning cDNA of an mRNA selected by the

-82-

preceding steps.

36. The method of claim 1 wherein the partitioning step is carried out after post-translationally modifying the nascent polypeptide.
- 5
37. The method of claim 1 wherein the partitioning step is carried out in the presence of a configurational modifier.
- 10
38. The method of claim 1 wherein step f) is carried out by chemical synthesis of the polypeptide ligand.
- 15
39. The method of claim 35 wherein step f) is carried out by expressing the cloned cDNA in a host cell.
- 20 40. The method of claim 1 wherein the mRNA additionally comprises a sequence encoding a C-terminal trailer.
- 25
41. The method of claim 1 wherein the mRNA additionally comprises a sequence encoding a segment of polypeptide that functions to bind a bridging molecule and step c) further comprises binding target molecules to a solid phase matrix and binding to the target molecules an anchor molecule covalently bound to the bridging molecule, the anchor molecule being capable of specifically binding the target molecules whereby ribosome
- 30

-83-

complexes bind to the bridging molecule anchored to the target molecules.

42. The method of claim 1 comprising the additional steps of g) synthesizing a second translatable mRNA mixture comprising the mRNA selected by steps a) - e) and a second randomized sequence coding region, and h) repeating steps b) - e) using the second translatable mRNA mixture to yield a desired level of target binding by a polypeptide encoded by the second mRNA enriched for sequences encoding the polypeptide.
- 15 43. A polypeptide ligand of a target molecule, said ligand made by the method of claim 1.
- 20 44. A polypeptide ligand of a protein, said ligand made by the method of claim 2.
45. A polypeptide ligand of an enzyme, said ligand made by the method of claim 3.
- 25 46. A polypeptide ligand of an antibody, said ligand made by the method of claim 4.
- 30 47. A polypeptide ligand of a receptor, said ligand made by the method of claim 5.
48. A polypeptide ligand of a nucleic acid binding protein, said ligand made by the

-84-

method of claim 6.

49. A polypeptide ligand of a toxin, said ligand made by the method of claim 7.
- 5 50. A polypeptide ligand of a glycoprotein, said ligand made by the method of claim 8.
- 10 51. A polypeptide ligand of an antigen, said ligand made by the method of claim 9.
52. A polypeptide ligand of a cell membrane component, said ligand made by the method of claim 11.
- 15 53. A polypeptide ligand of a virus component, said ligand made by the method of claim 12.
- 20 54. A polypeptide ligand of a carbohydrate, said ligand made by the method of claim 13.
- 25 55. A polypeptide ligand of a polysaccharide, said ligand made by the method of claim 14.
- 30 56. A polypeptide ligand of a lipid, said ligand made by the method of claim 15.
57. A polypeptide ligand of a glycolipid, said ligand made by the method of claim 16.

-85-

58. A polypeptide ligand of a toxin, said ligand made by the method of claim 17.
- 5 59. A polypeptide ligand of a drug, said ligand made by the method of claim 18.
60. A polypeptide ligand of a controlled substance, said ligand made by the method of claim 19.
- 10 61. A polypeptide ligand of a metabolite, said ligand made by the method of claim 20.
- 15 62. A polypeptide ligand of cofactor, said ligand made by the method of claim 21.
- 20 63. A polypeptide ligand of a nucleic acid, said ligand made by the method of claim 22.
64. A polypeptide ligand of a hormone, said ligand made by the method of claim 23.
- 25 65. A polypeptide ligand of a receptor ligand, said ligand made by the method of claim 24.
- 30 66. A polypeptide ligand of a transition state analog, said ligand made by the method of claim 25.
67. The method for selecting a polypeptide

-86-

ligand of a desired target molecule from a polypeptide mixture comprising:

5

- a) synthesizing a polypeptide mixture each member thereof having attached thereto amplifying means for separately amplifying the individual polypeptide to which it is attached;
- b) partitioning the polypeptide mixture with respect to binding the target molecule, thereby separating the mixture into polypeptide-target pairs and unbound polypeptides;
- c) amplifying the polypeptides of polypeptide-target pairs using said amplifying means; and
- d) repeating the partitioning and amplifying steps to select a polypeptide ligand of a desired target molecule.

10

15

20

68. The method of claim 67 wherein the polypeptide mixture comprises polypeptides having a segment of randomized amino acid sequence.

25

69. The method of claim 68 wherein the segment of randomized amino acid sequence is from 4 to 50 amino acids in length.

30

70. The method of claim 67 wherein the amplifying means comprises an mRNA mixture, each member thereof encoding a

-87-

polypeptide of the polypeptide mixture  
and being attached to the polypeptide it  
encodes as part of a ribosome complex.

- 5        71.      The method of claim 67 wherein the step  
                  of amplifying the polypeptides comprises  
                  the additional step of amplifying the  
                  mRNA mixture.
- 10       72.      The method of claim 71 wherein the mRNA  
                  mixture is amplified by reverse  
                  transcription and a polymerase chain  
                  reaction.
- 15       73.      A polypeptide ligand of a target  
                  molecule, said polypeptide being  
                  selected by the method of claim 64 and  
                  said target molecule being selected from  
                  the group consisting of a protein, an  
20                enzyme, a fusion protein, an antibody, a  
                  receptor, a receptor ligand, a nucleic  
                  acid, a nucleic acid binding protein, a  
                  glycoprotein, a toxin, an antigen, a  
                  cell membrane component, a virus, a  
25                virus component, a carbohydrate, a  
                  polysaccharide, a lipid, a glycolipid, a  
                  drug, a controlled substance, a hormone,  
                  a transition state analog, a metabolite  
                  or a cofactor.
- 30       74.      A method for selecting a nucleic acid  
                  comprising a sequence encoding a  
                  polypeptide ligand of a target molecule  
                  comprising the steps of:

-88-

- 5                   a) synthesizing a translatable mRNA mixture comprising a ribosome binding site, translation initiation codon and a randomized sequence coding region;
- 10                 b) synthesizing a mixture of ribosome complexes, each member thereof comprising a ribosome, a nascent polypeptide and a translated mRNA, said mRNA having a randomized coding region and said nascent polypeptide being the translation product of said mRNA;
- 15                 c) partitioning the ribosome complexes with respect to binding of the ribosome complex to a desired target molecule, thereby separating the ribosome complexes into ribosome complex-target pairs and unbound complexes, the ribosome complex-target pairs having mRNA enriched for sequences encoding target-binding polypeptides;
- 20                 d) amplifying the mRNA of partitioned ribosome complex-target pairs to yield a translatable mRNA mixture comprising a ribosome binding site, an initiation codon and a coding region enriched for sequences encoding target-binding polypeptides;
- 25                 e) repeating steps b) through d) using the mRNA enriched for sequences encoding target-binding
- 30

-89-

- polypeptides of each successive  
repeat as many times as desired to  
yield as desired level of target  
binding by a polypeptide encoded by  
the mRNA enriched for sequences  
encoding the polypeptide; and  
f) cloning a cDNA of an mRNA of step  
e), thereby selecting a nucleic  
acid comprising a sequence encoding  
a polypeptide ligand of a target  
molecule.

75. A polypeptide ligand of a target  
molecule, said polypeptide comprising a  
non-naturally-occurring sequence of  
amino acids, and being capable of  
binding a target molecule selected from  
the group consisting of a protein, an  
enzyme, a fusion protein, an antibody, a  
receptor, a nucleic acid, a nucleic acid  
binding protein, a glycoprotein, a  
toxin, an antigen, a cell membrane  
component, a virus, a virus component, a  
carbohydrate, a polysaccharide, a lipid,  
a glycolipid, a drug, a controlled  
substance, a hormone, a transition state  
analog, a metabolite or a cofactor.

76. A polypeptide having a minimum of about  
44 amino acids and a maximum of about  
100 amino acids, comprising of non-  
naturally-occurring binding segment of  
length ranging from about 4 to about 50  
amino acids and a C-terminal trailer

-90-

segment of about 40 amino acids, the binding segment having the property of binding to a target molecule.

- 5        77.      The polypeptide of claim 76 comprising additionally a N-terminal leader segment.
- 10        78..     The polypeptide of claim 76 wherein the binding segment comprises from about 10 to about 30 amino acids.
- 15        79.     The polypeptide of claim 76 wherein binding to the target molecule depends on the simultaneous presence of a configurational modifier.
- 20        80.     The polypeptide of claim 76 wherein binding to the target molecule depends on the presence of a disulfide bond in the polypeptide.
- 25        81.     A polypeptide comprising a binding segment of about 15-50 amino acids and having less than 50% amino acid homology with a naturally occurring amino acid sequence of the same length, the polypeptide having the property of binding to a desired target molecule.
- 30        82.     The polypeptide of claim 81 having less than 30% amino acid homology with a naturally occurring amino acid sequence of the same length.

-91-

83. The polypeptide of claim 81 further comprising a naturally-occurring sequence.

5

84. A method of modifying the function of a naturally-occurring protein comprising replacing a segment of said protein with a polypeptide made by the method of  
10 claim 1.

85. A protein modified by the method of  
claim 84 to contain a polypeptide having  
less than 50% homology with a naturally-  
15 occurring amino acid sequence of the  
same length.

86. A protein modified by the method of  
claim 84 to contain a polypeptide having  
less than 30% homology with a naturally-  
20 occurring amino acid sequence of the  
same length.

25

<sup>1/8</sup>  
SYSTEMATIC POLYPEPTIDE EVOLUTION BY REVERSE TRANSLATION.

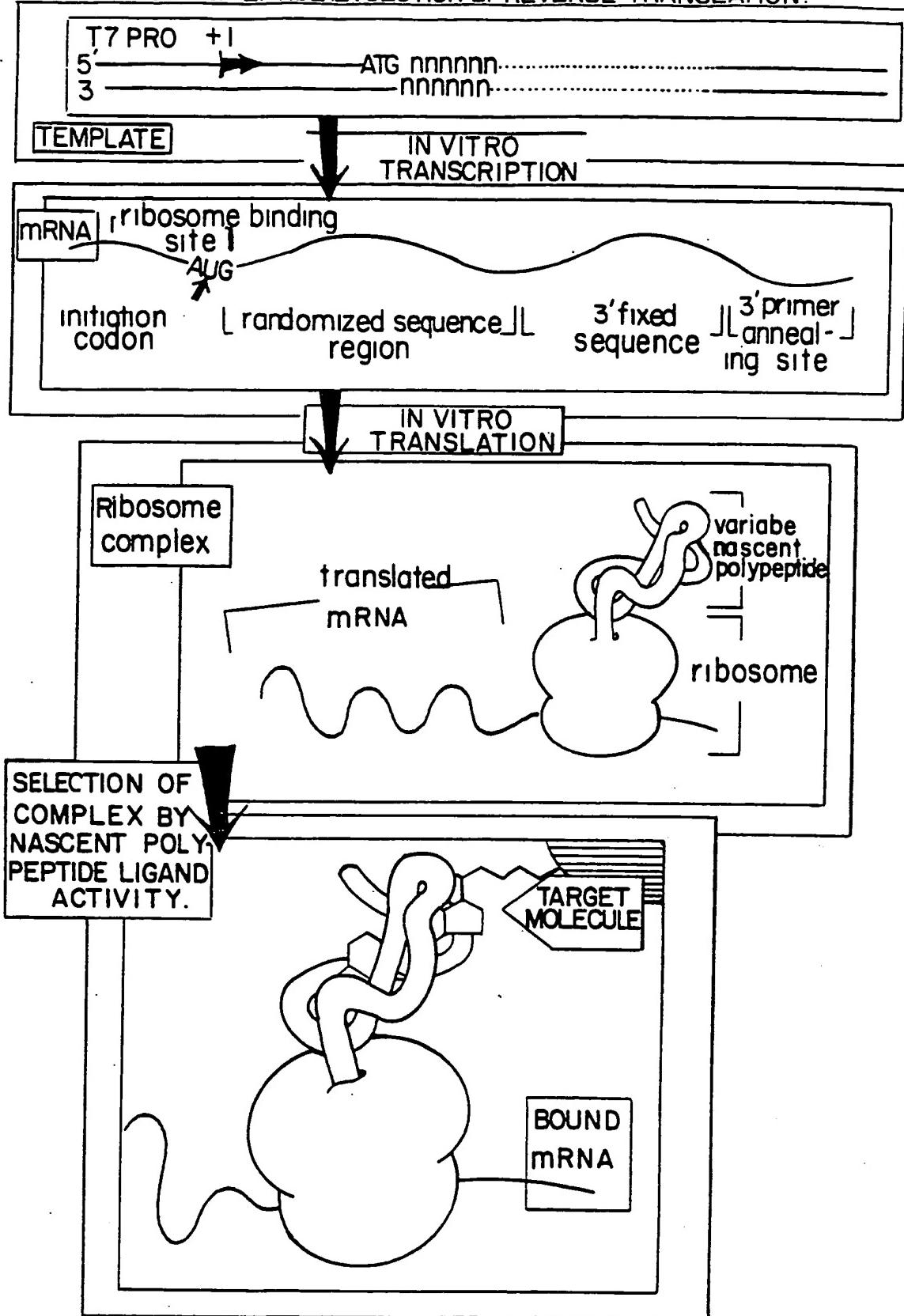


FIG. I

SUBSTITUTE SHEET

2/8.

## EXPANDED VIEWS OF RIBOSOME COMPLEX.

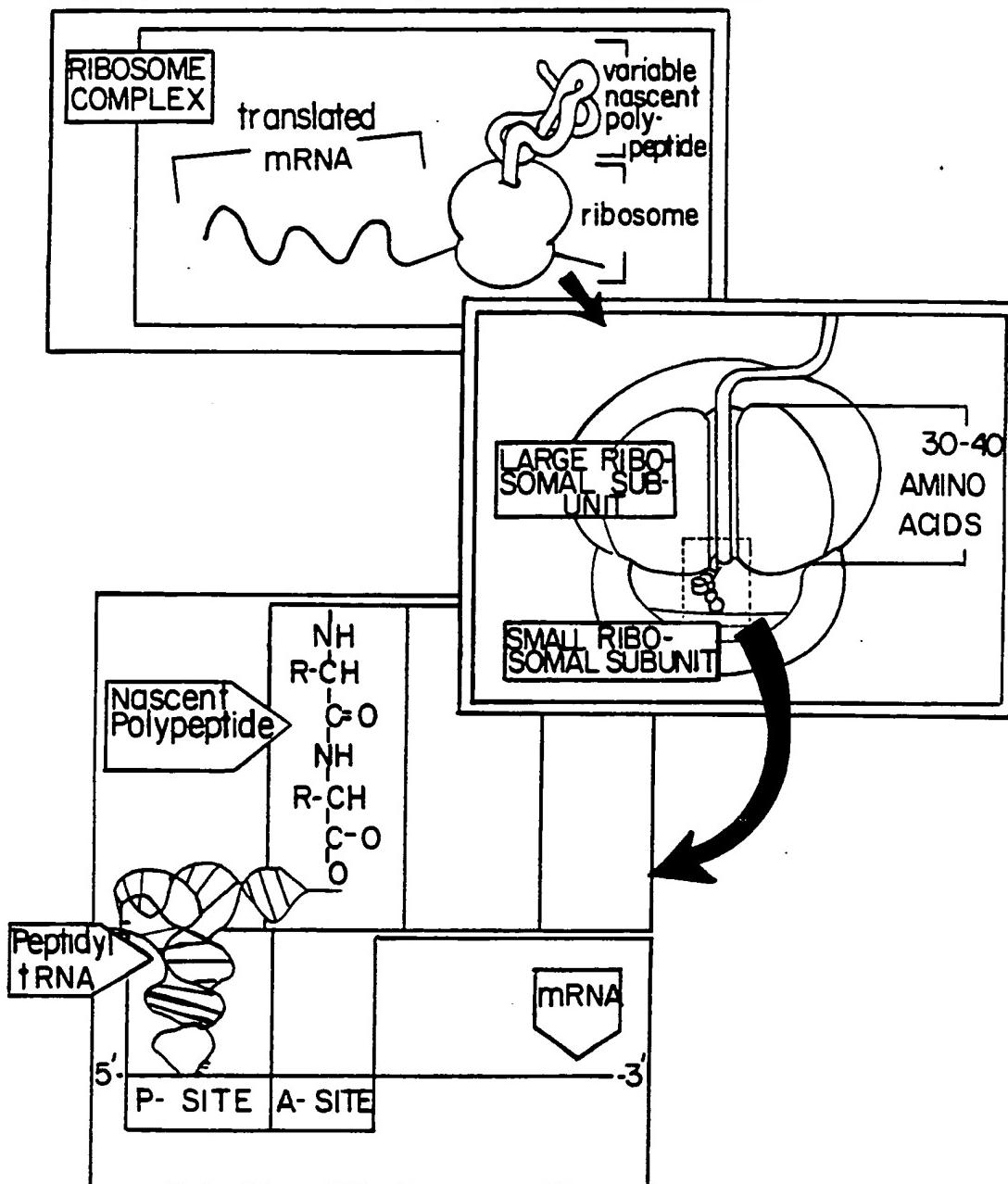


FIG. 2

SUBSTITUTE SHEET

3/8

## POLYPEPTIDE LIGANDS BY DIRECT IMMUNOPRECIPITATION

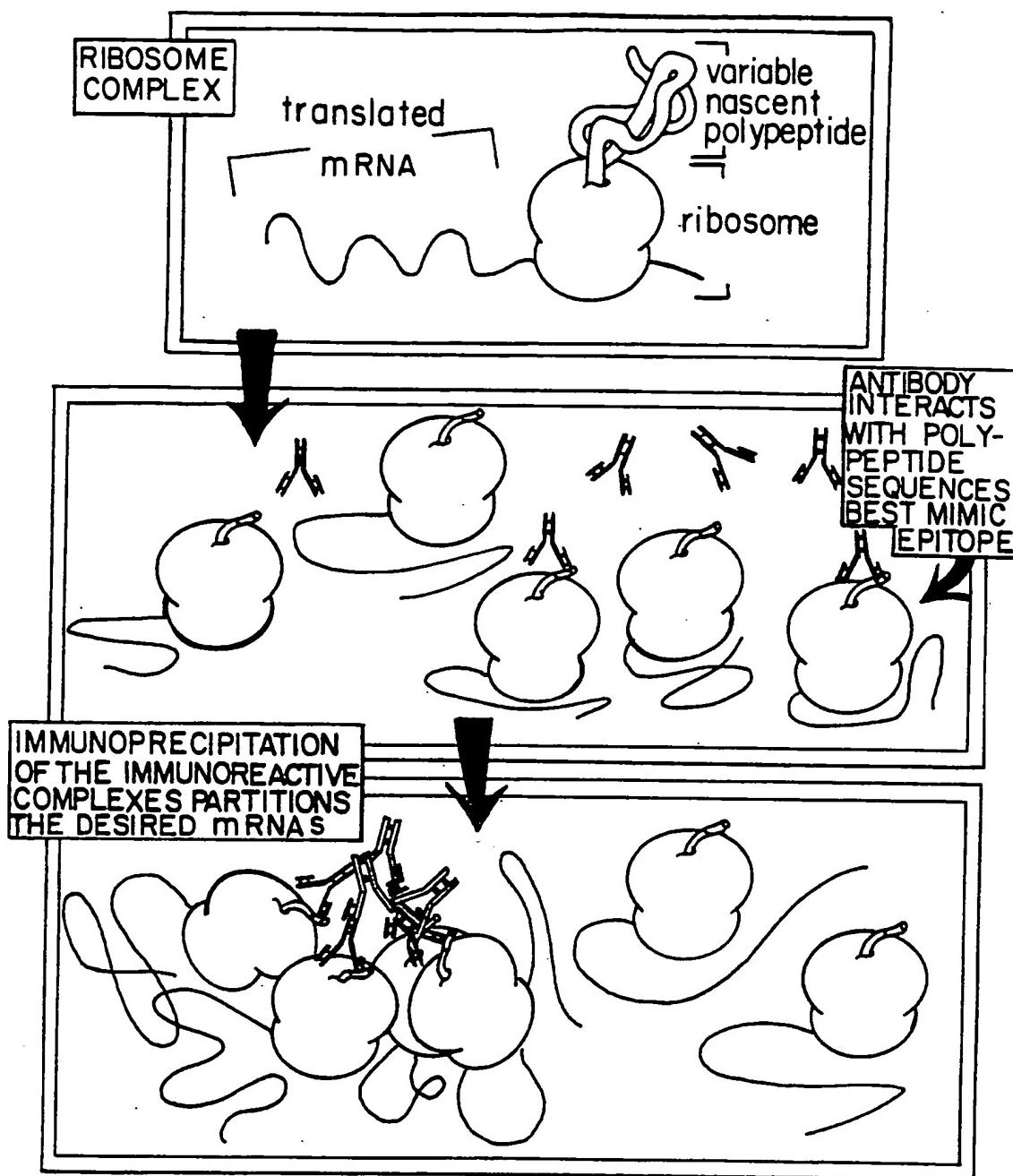


FIG. 3

SUBSTITUTE SHEET

4/8  
POLYPEPTIDE LIGANDS BY INDIRECT IMMUNOPRECIPITATION.

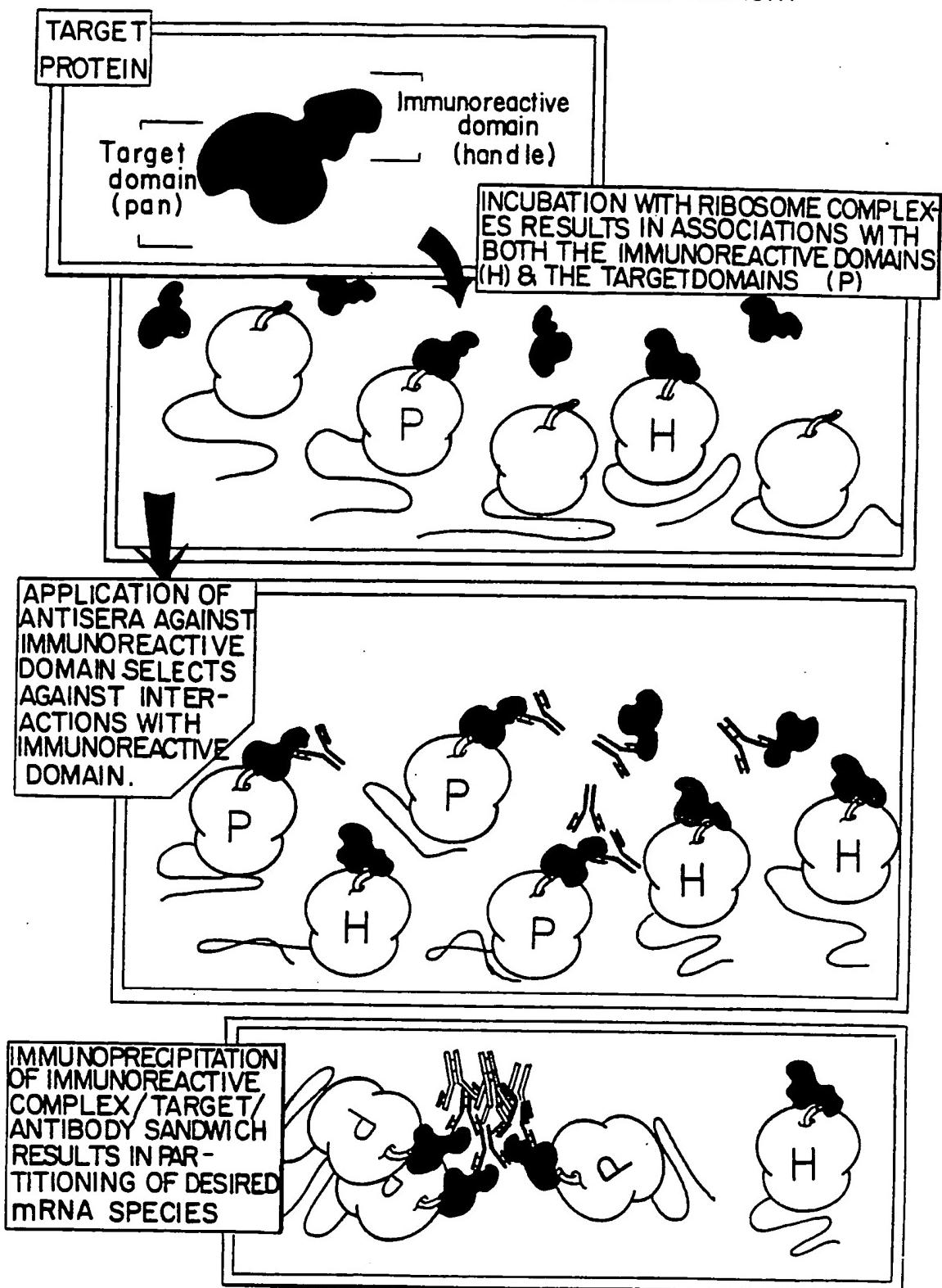


FIG. 4

5/8

## POLYPEPTIDE LIGANDS BY MEMBRANE PARTITIONING.

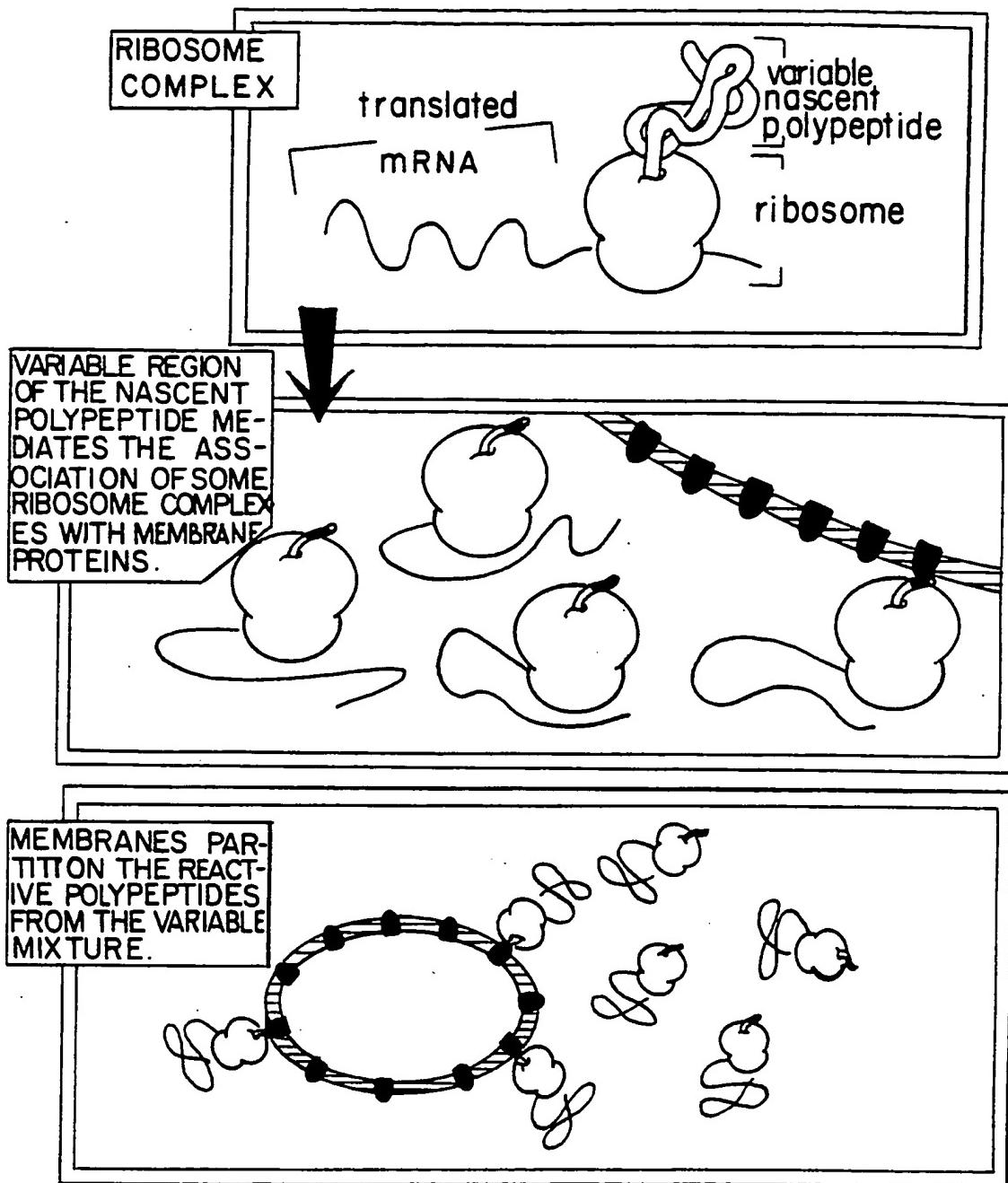


FIG. 5

6/8  
POLYPEPTIDE LIGANDS BY AFFINITY COLUMN CHROMATOGRAPHY.

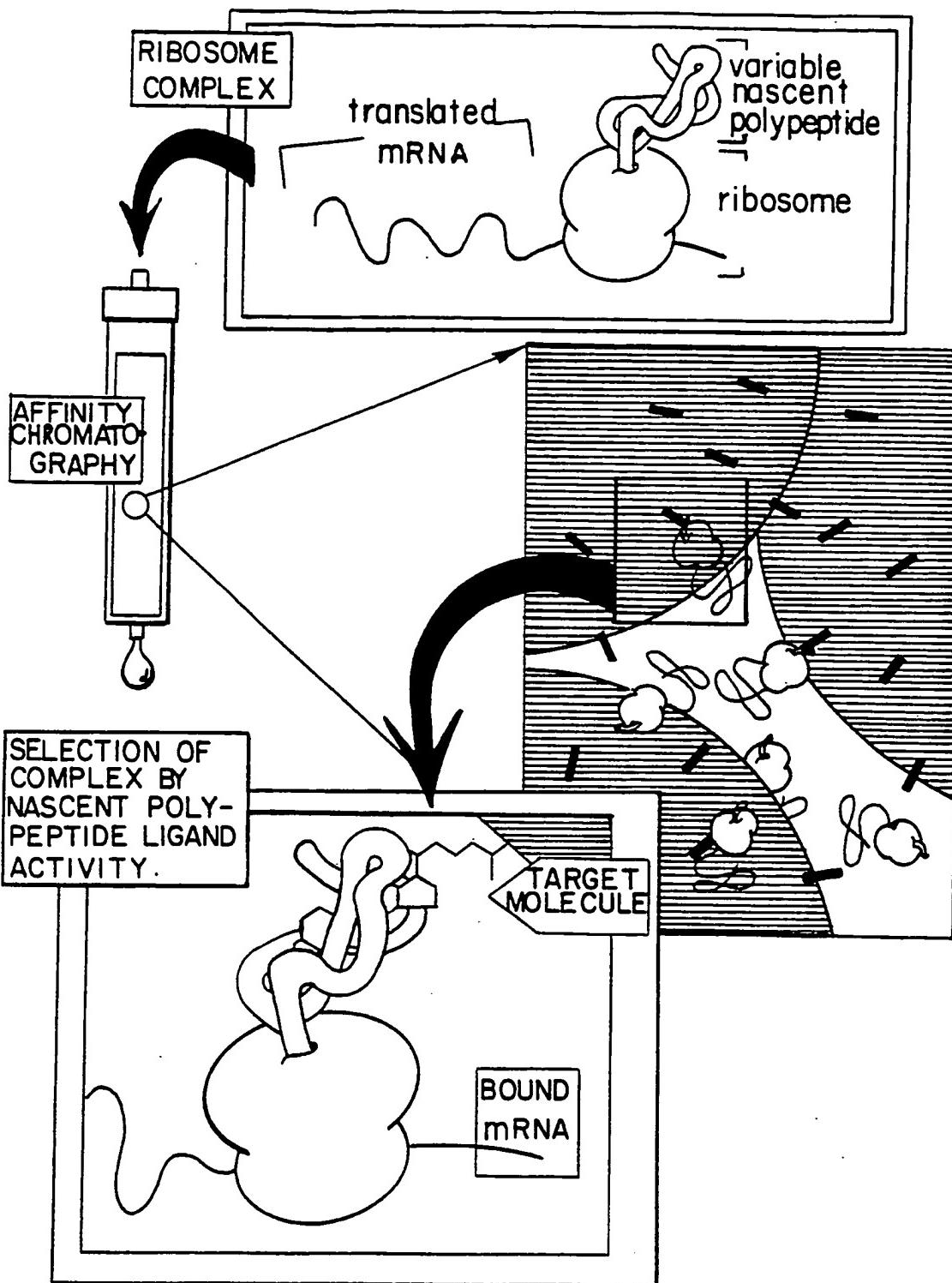


FIG. 6

7/8

ANCHORING OF BRIDGING EPITOPE & SECONDARY LIGAND  
EVOLUTION.

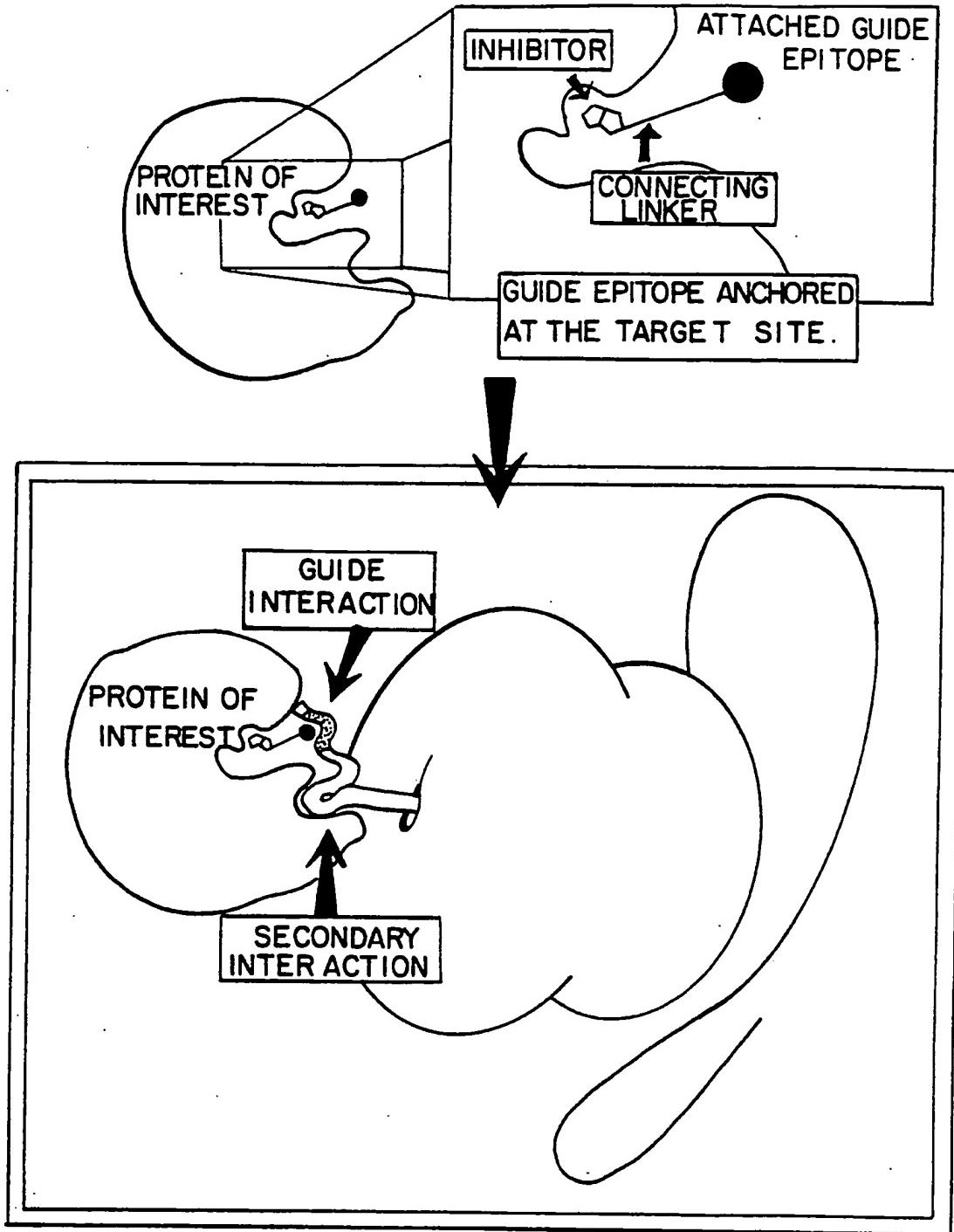


FIG. 7

SUBSTITUTE SHEET

8/8

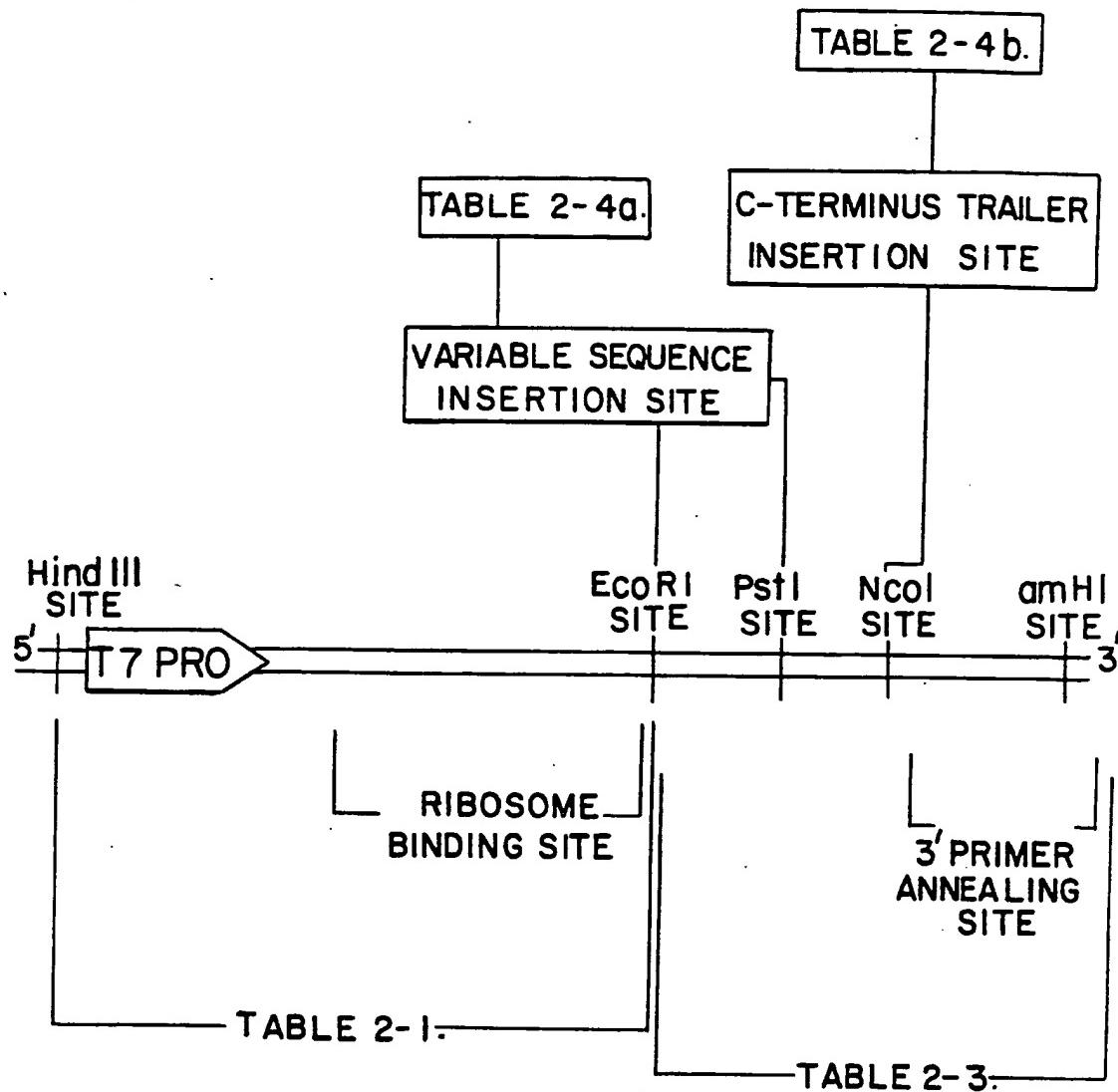


FIG. 8

# INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/05463

## I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) \*

According to International Patent Classification (IPC) or to both National Classification and IPC  
 IPC(S): C07K 3/00;C12N15/00  
 U.S.CI. 530/350;435/172.3

## II. FIELDS SEARCHED

### Minimum Documentation Searched 7

Classification System	Classification Symbols
U.S.CI.	530/350 ; 435/172.3

Documentation Searched other than Minimum Documentation  
 to the Extent that such Documents are Included in the Fields Searched 8

APS, Dialog Files, 5,155,73,76

## III. DOCUMENTS CONSIDERED TO BE RELEVANT 9

Category *	Citation of Document, <sup>11</sup> with indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>
Y	Abstract. meeting on RNA Proceeding; May 16-May 20. 1990. Ellington et al. "Selection of RNAs with Ligand-specific Binding Activity from Pools of Random sequence molecules", pages 224-226, see entire article.	1-3, 26-44, 67-83
Y	Science. Volume 249. issued 27 July 1990 Devlin et al. "Random peptide libraries: A source of specific protein binding molecules". pages 404-406, see entire article.	1-3, 26-44, 67-83

\* Special categories of cited documents: 10

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"Z" document member of the same patent family

## IV. CERTIFICATION

Date of the Actual Completion of the International Search

19 November 1991

Date of Mailing of this International Search Report

18 DEC 1991

International Searching Authority

ISA/US

Signature of Authorized Officer

H. Sidberry

G. Morris Jr.

**III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)**

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	Science, volume 249, issued 27 JULY 1990, Scott et al., "Searching for Peptide Ligands with an Epitope Library." pages 386-390, see entire article.	1,2,26-44, 67-83
Y,P	Science, volume 249, issued 03 AUGUST 1990, Tuerk et. al., "Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase", pages 505-510, see entire article.	1,2,26-44, 67-83